



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 122955

TO: Elizabeth Kemmerer

Location: 09/937905

Art Unit: 1646

June *2*, 2004 *4090*

Case Serial Number: 09/937905

From: P. Sheppard

Location: Remsen Building

Phone: (571) 272-2529

sheppard@uspto.gov

Search Notes

371 - 3/31/2000
JP - 4/1/99

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 1, 2004, 19:32:41 ; Search time 5516 Seconds
(without alignments)
6600.462 Million cell updates/sec

Title: US-09-937-905-1
Perfect score: 840
Sequence: 1 gaaccatgtctggtgtca.....aaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	840	100.0	840	6	BD107339	BD107339 Novel cho
2	840	100.0	840	6	BD168026	BD168026 Novel pro
3	840	100.0	840	10	AB028863	AB028863 Mus muscu
4	823.2	98.0	879	10	BC028434	BC028434 Mus muscu
5	598.6	71.3	1183	6	AX235607	AX235607 Sequence
6	597	71.1	1132	9	BC009077	BC009077 Homo sapi
7	595.4	70.9	1135	9	BC017594	BC017594 Homo sapi
8	595.4	70.9	1136	6	BD168027	BD168027 Novel pro
9	595.4	70.9	1190	6	E28764	E28764 HFI2G53 pol
10	593.8	70.7	1199	9	BC008440	BC008440 Homo sapi
11	593.8	70.7	1226	6	BD107338	BD107338 Novel cho
12	593.8	70.7	1226	6	AF132963	AF132963 Homo sapi
13	576.2	68.6	103861	9	AL807761	AL807761 Human DNA
14	576.2	68.6	147844	2	AC068188	AC068188 Homo sapi
15	576.2	68.6	192001	2	AC068376	AC068376 Homo sapi
16	484.6	57.7	3337	9	BC010133	BC010133 Homo sapi
17	474.2	48.5	834	6	E28765	E28765 HFI2G53 pol
18	360	42.9	758	6	BD216184	BD216184 Novel hum
19	354	42.1	778	6	BD216494	BD216494 Novel hum
20	348.2	41.5	851	6	BD215002	BD215002 Novel hum
21	323	38.5	409	6	AR417783	AR417783 Sequence
22	323	38.5	409	6	BD113336	BD113336 EST and e
23	262.8	31.3	913	3	AY071553	AY071553 Drosophil
24	248.8	29.6	300	6	BD212080	BD212080 Novel hum
25	245.6	29.2	300	6	BD212079	BD212079 Novel hum
26	218	26.0	1538	3	AK113469	AK113469 Clona int
27	187.8	22.4	642	9	HUMZC34F03	AF088034 Homo sapi
28	178.4	21.2	158450	10	AC068911	AC068911 Mus muscu
29	178.4	21.2	163615	10	AC068913	AC068913 Mus muscu
30	178.4	21.2	233417	10	AL844146	AL844146 Mouse DNA
31	177.8	21.2	32746	2	AC013224	AC013224 Drosophil
32	177.8	21.2	168384	3	AE003493	AE003493 Drosophil
33	177.8	21.2	307761	3	AE003493	AE003493 Rattus no
34	161.2	19.2	250865	2	AC094899	AC094899 Rattus no
35	161.2	19.2	270241	2	AC112064	AC112064 Rattus no
36	161.2	19.2	298446	2	AC117043	AC117043 Rattus no
37	145	17.3	160764	9	AC107928	AC107928 Homo sapi
38	145	17.3	176593	2	AC087661	AC087661 Homo sapi
39	145	17.3	179901	2	AC011006	AC011006 Homo sapi
40	143	17.0	229	6	AR048129	AR048129 Sequence
41	143	17.0	229	6	AR087771	AR087771 Sequence
42	143	17.0	229	6	AR170853	AR170853 Sequence
43	138	16.4	1852	8	AK065321	AK065321 Oryza sat
44	121.8	14.5	121027	2	AL157891	AL157891 Homo sapi
45	121.8	14.5	128228	9	AL138810	AL138810 Human DNA

ALIGNMENTS

RESULT 1	BD107339	840 bp	DNA	linear	PAT 18-SEP-2002
LOCUS	BD107339				
DEFINITION	Novel chondrogenesis promoter.				
ACCESSION	BD107339				
VERSION	BD107339.1	GI:23202157			
KEYWORDS	JP 2002020311-A/2.				
SOURCE	Mus sp.				
ORGANISM	Mus sp.				
REFERENCE	1 (bases 1 to 840)				
AUTHORS	Kato, Y. and Fujimoto, K.				
TITLE	Novel chondrogenesis promoter				
JOURNAL	Patent: JP 2002020311-A 2 23-JAN-2002;				

```
COMMENT
JAPAN SCIENCE AND TECHNOLOGY CORP
OS Mus sp. (mouse)
PN JP 20020203111-A/2
PD 23-JAN-2002
PI 07-JUL-2000 JP 2000206566
PY YUKIO KATO, KATSUMI FUJIMOTO
PC A61K38/00, A61P19/00, A61P19/02, A61P19/08, C07K16/18, C12N15/09,
C12Q1/68,
PC G01N33/53, G01N33/566, C07K14/42, C07K14/51, C07K14/79, A61K37/02,
C12N15/00
PC Novel chondrogenesis promoter
CC Location/Qualifiers
FH Key 1..840
FT source /organism="Mus sp. (mouse)".
FEATURES
source
Location/Qualifiers
1..840
/organism="Mus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10095"
ORIGIN
Query Match 100.0%; Score 840; DB 6; Length 840;
Best Local Similarity 100.0%; Pred. No. 5.1e-198;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAACCATGTCGTGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCCGCTGTGGCGCGCAGG 60
DB 1 GAACCATGTCGTGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCCGCTGTGGCGCGCAGG 60
QY 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGCAACAGTTTACCATCTGGGCT 120
DB 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGCAACAGTTTACCATCTGGGCT 120
QY 121 GGGTCATGGCAGCTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGTC 180
DB 121 GGGTCATGGCAGCTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGTC 180
QY 181 CCTCAGGCGTCAAAAGAGGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
DB 181 CCTCAGGCGTCAAAAGAGGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAGAGCCAGTGCACTCCTC 300
DB 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAGAGCCAGTGCACTCCTC 300
QY 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTTAAG 360
DB 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTTAAG 360
QY 361 CTGCTGTGATGCTACCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
DB 361 CTGCTGTGATGCTACCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
QY 421 TGATCAAGAGTAAGGAAATGTACCTCAGAGGCTTTACAGATACCATGATATGTTAG 480
DB 421 TGATCAAGAGTAAGGAAATGTACCTCAGAGGCTTTACAGATACCATGATATGTTAG 480
QY 481 TGGTACCTATTATTAGAGAACCTCTCTGAAGAGAGGATCTCAAGAGAGGATGGCTCATG 540
DB 481 TGGTACCTATTATTAGAGAACCTCTCTGAAGAGAGGATCTCAAGAGAGGATGGCTCATG 540
QY 541 CCATGAATGAGTACCCAGACTCCTGTGGGTTCTTGTCCGGCGTCATGGGTTGACGTGT 600
DB 541 CCATGAATGAGTACCCAGACTCCTGTGGGTTCTTGTCCGGCGTCATGGGTTGACGTGT 600
QY 601 GGGGAGAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
DB 601 GGGGAGAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
QY 661 TTGCTGTCTCCATGAGAGAGATGGGACTCGATCCACACAGCTCCAGTTGGAGAAATG 720
DB 661 TTGCTGTCTCCATGAGAGAGATGGGACTCGATCCACACAGCTCCAGTTGGAGAAATG 720
```

```
QY 721 GAATTGTGTAAGCCAAGTGGATCCCTAAGCATCTCCAAACAATAAAAACTCAATTATG 780
DB 721 GAATTGTGTAAGCCAAGTGGATCCCTAAGCATCTCCAAACAATAAAAACTCAATTATG 780
QY 781 CCTTAATAAACTCAGCTGCTTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 840
DB 781 CCTTAATAAACTCAGCTGCTTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 840
RESULT 2
BD168026 840 bp DNA linear PAT 17-JAN-2003
LOCUS Novel proteins, gene encoding the same and method of utilization
DEFINITION thereof.
ACCESSION BD168026
VERSION BD168026.1 GI:27873838
KEYWORDS WO 0226978-A/1.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE
1 (bases 1 to 840)
AUTHORS Sha, S., Mukai, H., Aoki, Y. and Nishi, Y.
TITLE Novel proteins, gene encoding the same and method of utilization
JOURNAL Patent: WO 0226978-A 1 04-APR-2002;
JAPAN TOBACCO INC., SHIKEN SHA, HIDEHITO MUKAI, YOSHIKO AOKI, YOSHISUKE
NISHI
COMMENT
OS Mouse macrophage cell RAW 264.7
PN WO 0226978-A/1
PD 04-APR-2002
PF 27-SEP-2001 WO 2001JP008446
PR 27-SEP-2000 JP 00P 294191
PI SHIKEN SHA, HIDEHITO MUKAI, YOSHIKO AOKI, YOSHISUKE NISHI PC
C12N15/12, C12N1/21, C07K14/705, C07K16/28, A61K38/00, A61K39/395, PC
A61K45/00,
PC A61K48/00, A61P7/00, A61P43/00, G01N33/15, G01N33/50, PC
G01N33/566//
PC C12P21/08, (C12N15/12, C12R1:91), (C12N1/21, C12R1:19) CC Novel
proteins, gene encoding the same and method of CC
utilization thereof
FH Key Location/Qualifiers
FT source 1..840
/organism="Mouse macrophage cell RAW 264.7".
FEATURES
source
Location/Qualifiers
1..840
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 100.0%; Score 840; DB 6; Length 840;
Best Local Similarity 100.0%; Pred. No. 5.1e-198;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAACCATGTCGTGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCCGCTGTGGCGCGCAGG 60
DB 1 GAACCATGTCGTGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCCGCTGTGGCGCGCAGG 60
QY 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGCAACAGTTTACCATCTGGGCT 120
DB 61 ACAAGGAGACCCCGGATTCCTGATCCGAGAACTTTGCAACAGTTTACCATCTGGGCT 120
QY 121 GGGTCATGGCAGCTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGTC 180
DB 121 GGGTCATGGCAGCTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCGTC 180
QY 181 CCTCAGGCGTCAAAAGAGGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
DB 181 CCTCAGGCGTCAAAAGAGGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAGAGCCAGTGCACTCCTC 300
DB 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAGAGCCAGTGCACTCCTC 300
```

```
301 TTTTCATGAATGCTTATACCATGAGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Db TTTTCATGAATGCTTATACCATGAGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Qy CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCATCAAGAGA 420
Db CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCATCAAGAGA 420
Qy TGATCAAAAGGAATTAAGGAATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
Db TGATCAAAAGGAATTAAGGAATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
Qy TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGAGATCTCAAGAAAGGATGCTCATG 540
Db TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGAGATCTCAAGAAAGGATGCTCATG 540
Qy CCATGAATGATGCCAGACCTCTGCGGTCTTGTCCGGCGTCATGGGGTGTACGTGT 600
Db CCATGAATGATGCCAGACCTCTGCGGTCTTGTCCGGCGTCATGGGGTGTACGTGT 600
Qy GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Db GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Qy TTGCTGTCTCCATGAAGAAGATGGACCTCGATCCAAACACAGCTCCAGTCTGGAGAAATG 720
Db TTGCTGTCTCCATGAAGAAGATGGACCTCGATCCAAACACAGCTCCAGTCTGGAGAAATG 720
Qy GAATTTGTGTAAGCCAGTGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 780
Db GAATTTGTGTAAGCCAGTGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 780
Qy CCTTAAATTAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 840
Db CCTTAAATTAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 840

RESULT 3
AB028863 Mus musculus mRNA for MMRP19, complete cds. linear ROD 17-JUN-1999
LOCUS AB028863
DEFINITION Mus musculus mRNA for MMRP19, complete cds.
ACCESSION AB028863
VERSION AB028863.1 GI:5103282
KEYWORDS MMRP19.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Shi, S., Aoki, Y. and Nishi, Y.
A cDNA sequence from murine monocyte-macrophage
Published Only in DataBase (1999)
2 (bases 1 to 840)
Shi, S., Aoki, Y. and Nishi, Y.
Direct Submission
Submitted (10-JUN-1999) Shiken Sha, JAPAN TOBACCO INC., Laboratory
of Life Science & Biomolecular Engineering; 6-2, Umegaoka, Aoba-ku,
Yokohama, Kanagawa 227-8512, Japan
(E-mail: shiken_sha@ims.jti.co.jp, Tel: 81-45-972-5901,
Fax: 81-45-972-6205)
Location/Qualifiers
1..840
/organism="Mus musculus"
/mol_type="mRNA"
/sub_species="scares mouse"
/db_xref="taxon:10090"
/clone="MMR19"
/sex="male"
/cell_line="RAW 264.7"
/cell_type="Monocyte-macrophage"
/clone_lib="a cDNA library of RAW 264.7 cell"
1..840
/gene="MMR19"

gene
```

CDS

```
6..731
/gene="MMR19"
/codon_start=1
/product="MMRP19"
/protein_id="BAA78906.1"
/db_xref="GI:5103283"
/translation="MSGCAAGGCCPPCGAOKDEHPRELIPELKOFYHLGWVTGTG
GGLSLKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLKKSQCTPLEWN
AYTMRGAGAVITHSKAAVNATLLFFQGFKITHQEWIKRKIKTSKSGYTRYDDMLVV
PIITNPEEKDLKRMAMHANEYDPSCAVLVRHGVYVWGETWEKAKTMCCECYDLED
IAVSMKKMGLDPTQLPVGENGIY"

ORIGIN
Query Match 100.0%; Score 840; DB 10; Length 840;
Best Local Similarity 100.0%; Pred. No. 5.1e-198; Indels 0; Gaps 0;
Matches 840; Conservative 0; Mismatches 0;

QY 1 GAACCATGTCTGGCTGTCAAGGCTCAAGGAGACTGTTGCTCGCGCCCTGTGGCGCGCAGG 60
Db 1 GAACCATGTCTGGCTGTCAAGGCTCAAGGAGACTGTTGCTCGCGCCCTGTGGCGCGCAGG 60
QY 61 ACAAGGAGCACCCCGATTCCTGATCCAGAACCTTTGCAAAACAGTTTACCATCTGGCT 120
Db 61 ACAAGGAGCACCCCGATTCCTGATCCAGAACCTTTGCAAAACAGTTTACCATCTGGCT 120
QY 121 GGGTCATCGCAGCTGGAGGGGAGTCACTTGAGCATGGCAATGCAATCTACATTGCTC 180
Db 121 GGGTCATCGCAGCTGGAGGGGAGTCACTTGAGCATGGCAATGCAATCTACATTGCTC 180
QY 181 CCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTATG 240
Db 181 CCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTATG 240
QY 241 AGCAGGACATAAGCGGCTCTCAGCATCTAAGAGCTGAAAAAGCAGTGCTCCTC 300
Db 241 AGCAGGACATAAGCGGCTCTCAGCATCTAAGAGCTGAAAAAGCAGTGCTCCTC 300
QY 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
Db 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGCGCAGTGATTCATACCCACTCTAAAG 360
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420
Db 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420
QY 421 TGATCAAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
Db 421 TGATCAAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTTAG 480
QY 481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGATCTCAAGAAAGGATGCTCATG 540
Db 481 TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGATCTCAAGAAAGGATGCTCATG 540
QY 541 CCATGAATGATGCCAGACCTCTGCGGTCTTGTCCGGCGTCATGGGGTGTACGTGT 600
Db 541 CCATGAATGATGCCAGACCTCTGCGGTCTTGTCCGGCGTCATGGGGTGTACGTGT 600
QY 601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
Db 601 GGGGAGAAACATGGGAGAAACCAAAACCATGTGTGAGTGTATTGACTTACCTGTTTGACA 660
QY 661 TTGCTGTCTCCATGAAGAAGATGGACCTCGATCCAAACACAGCTCCAGTCTGGAGAAATG 720
Db 661 TTGCTGTCTCCATGAAGAAGATGGACCTCGATCCAAACACAGCTCCAGTCTGGAGAAATG 720
QY 721 GAATTTGTGTAAGCCAGTGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 780
Db 721 GAATTTGTGTAAGCCAGTGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 780
QY 781 CCTTAAATTAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 840
Db 781 CCTTAAATTAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAGGATGCTTAAAG 840
```


RESULT 4

BC028434

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

gene

BC028434 879 bp mRNA linear ROD 07-OCT-2003
 IMAGE:1245515), complete cds.
 BC028434
 BC028434.1 GI:20306407
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 879)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.J., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 879)
 Strausberg, R.
 Direct Submission
 Submitted (23-APR-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbe-r@mail.nih.gov
 Tissue Procurement: Marcello Bento Soares, Ph.D.
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 66 Row: 9 Column: 16
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 9790106.
 Location/Qualifiers
 1. .879
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="MGC:41093 IMAGE:1245515"
 /tissue_type="Mammary gland"
 /clone_lib="Soares_mammary_gland_NbMMG"
 /lab_host="DH10B"
 /note="Vector: p17T3-Pac"
 1. .879

/gene="Mmrp19-pending"
 /note="synonym: Mmrp19"
 /db_xref="LocusID:56369"
 /db_xref="MGI:1926788"
 57. .782
 /codon_start=1
 /product="cDNA sequence AB028863"
 /protein_id="AAH28434.1"
 /db_xref="GI:20306408"
 /db_xref="LocusID:56369"
 /translation="MSGCOAGDCCSRPCGADKEHPRELIPELCKOFVHLGWVTGTG
 GGSILKHGNEIYIAPSGVKERIQPEDMFVCDINODISEPPASKKLKSOCTPLPMN
 AITMKGAVIHTSKAAVMATLLFPQEFKITHQEMINGIRKCTSGGYYRVDMLVY
 PIITMPEKLDKERMAHMYDPSCAVLVRHHGVYVWGETWEKAKTWCECYDYLFD
 IAVSMKKMDLPTQLPVGENGIV"
 132. .758
 misc_feature
 /note="Aldolase II; Region: Class II Aldolase and Adducin
 N-terminal domain. This family includes class II aldolases
 and adducins which have not been ascribed any enzymatic
 function"
 /db_xref="CDD:pfam00596"
 ORIGIN
 Query Match 98.0%; Score 823.2; DB 10; Length 879;
 Best Local Similarity 99.8%; Pred. No. 7.6e-194;
 Matches 825; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GAACCATGTCTGGCTCTCAAGCTCAAGGAGACTGTGTCTCGCGCGCTGTGGCGCGCAGG 60
 DB 52 GAACCATGTCTGGCTCTCAAGCTCAAGGAGACTGTGTCTCGCGCGCTGTGGCGCGCAGG 111
 QY 61 ACAAGAGACACCCCGATTCCTGATCCAGACTTTGGAAAAGATTGACATCTGGGCT 120
 DB 112 ACAAGAGACACCCCGATTCCTGATCCAGACTTTGGAAAAGATTGACATCTGGGCT 171
 QY 121 GGTCTACTGCGCTGAGGGGGGAAATCAGCTTGAAGCATGCAATGAATCTACATGCTC 180
 DB 172 GGTCTACCGGCACCTGGAGGGGGAATCAGCTTGAAGCATGCAATGAATCTACATGCTC 231
 QY 181 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGACATTAATG 240
 DB 232 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGACATTAATG 291
 QY 241 AGCAGACATAAGCGGGGCTTCCAGCATCTAAGAAAGCTGAAAAGAGCAGTGCATCTCCT 300
 DB 292 AGCAGACATAAGCGGGGCTTCCAGCATCTAAGAAAGCTGAAAAGAGCAGTGCATCTCCT 351
 QY 301 TTTTTCATGATGCTTATACCATGAGAGAGCTGCGCAGTGATTTCATCCCATCTCTAAAG 360
 DB 352 TTTTTCATGATGCTTATACCATGAGAGAGCTGCGCAGTGATTTCATCCCATCTCTAAAG 411
 QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACACATCAAGAGA 420
 DB 412 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACACATCAAGAGA 471
 QY 421 TGATCAAGAGATAAGAAATGTACCTCAGGAGCTTATACAGATACGATGATGTAG 480
 DB 472 TGATCAAGAGATAAGAAATGTACCTCAGGAGCTTATACAGATACGATGATGTAG 531
 QY 481 TGTGTACTATTATTGAGAACACCTCTGAGAGAGGATCTCAAGAGAAAGATGGCTCATG 540
 DB 532 TGTGTACTATTATTGAGAACACCTCTGAGAGAGGATCTCAAGAGAAAGATGGCTCATG 591
 QY 541 CCAATGAATGAGTACCCAGACTCTCTGTGCGGTTCTTGTGCGGCTCATGCGGTTGACGTGT 600
 DB 592 CCAATGAATGAGTACCCAGACTCTCTGTGCGGTTCTTGTGCGGCTCATGCGGTTGACGTGT 651
 QY 501 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACTCTGTTGACA 660
 DB 652 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACTCTGTTGACA 711
 QY 661 TTGCTGTCTCCATGAAGAAGATGGGACTCCGATCCAAACACAGCTCCCAAGTTGGAGAAATG 720

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: ang@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.B., Sheed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 13 Row: P Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

Location/Qualifiers

1..1132

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:9397 IMAGE:3875061"

/tissue_type="Eye, retinoblastoma"

/clone_lib="NIH_MGC_67"

/lab_host="DH10B"

/notes="vector: pCMV-SPORT6"

1..1132

/genes="MMRP19"

/db_xref="LocusID:51074"

42..770

/codon_start=1

/product="CGI-29 protein"

/protein_id="AAH09077.1"

/db_xref="GI:14290597"

/db_xref="LocusID:51074"

/translation="MSGCDAREGDCCSRRCQADKEHPYVLIPELCKQFVHLGWVGT
GGISLKHGEIYIARSGVOKERLOPEDMEVCDINEKDLSGSPSKLKSQCTPLFM
NAYTMKGAVIHTIARSGVOKERLOPEDMEVCDINEKDLSGSPSKLKSQCTPLFM
VPIENTPEEKIDKRAHAMNPEYDSCAVLVRHGVYWGVTWEKAKTMCEYDYL
DIASVKKVGLDPSQLPVGNGIV"

117..746

/notes="Aldolase II; Region: Class II Aldolase and Adducin
N-terminal domain. This family includes class II aldolases
and adducins which have not been ascribed any enzymatic
function"

/db_xref="CDD:pfam00596"

ORIGIN

Query Match 71.1%; Score 597; DB 9; Length 1132;
Best Local Similarity 84.8%; Pred. No. 1.3e-137;
Matches 693; Conservative 0; Mismatches 120; Indels 4; Gaps 2;
QY 4 CCATGCTGGCTGTCAAGCTC---AAGGAGACTGTGTCGGCGCCCTGTGGCGCGCAGG 60
DB 40 CCATGCTGGCTGTGATGCTCGGAGGAGACTGTGTCGGGAGATGGCGCGCAGG 99
QY 61 ACAAGGAGACCCCGGATTCCTGATCCAGACTTTCCAGACTTTTCAATCCATCGGCT 120
DB 100 ACAAGGAGACATCAAGATACCTGATCCAGACTTTCCAGACTTTTCAATCCATCGGCT 159
QY 121 GGGTCACTGGCACTGGAGGGGATCAGCTTGAGCATGGCAATGAATCTACATTCCTC 180
DB 160 GGGTCACTGGCACTGGAGGGGATCAGCTTGAGCATGGCAATGAATCTACATTCCTC 219
QY 181 CCTCAGGCGTCAAAAGAGCGCATTCAGCCAGAGACATGTTTGTGTGTGACATTAATG 240
DB 220 CTTCAGAGTCAAAAGAGCGAATTCAGCTTGAGCATGTTTGTGTGTGATATAATG 279
QY 241 AGCAGGACATAGCGGCGCTCCAGCATCTTAAGAGCTGAAAAAGCCAGTGCATCCTC 300
DB 280 AAAAGGACATAGTGGACCTTCGCCATCGAAGAGCTAAAAAGCCAGTGTACTCCTC 339
QY 301 TTTTTCATGAATGCTTATACCATGAGAGAGCTGGCGCAGTGAATTCATCCCACTTAAG 360
|||||

Db 340 TTTTTCATGAATGCTTACACAATGAGAGGAGCAGGTGCAGTGATTCATACCCACTCTAAAG 399
QY 361 CTGCTGTGATGGTACCCCTTCTGTTTCCAGGACAGGAGTTTAAATTTACACATCAAGAGA 420
Db 400 CTGCTGTGATGGCCACACTTCTTTCAGAGCGGAGTTTAAATTTACACATCAAGAGA 459
QY 421 TGATCAAAAGGAATAAGAAATGTACTTCCGAGGGGTATTATAGATATGATATGTTAG 480
Db 460 TGATAAAGGAATAAGAAATGTACTTCCGAGGGGTATTATAGATATGATATGTTAG 519
QY 481 TGGTACCTTATTGAGACACCTCTCGAGAGAGGAGTCTCAAGAGAAAGGATGGCTCATG 540
Db 520 TGTATCCCATTTATGAGATAACCTCGAGAGAGAGACCTCAAGATAGATAGATGGCTCATG 579
QY 541 CCATGAATGATGATCCAGACTCTCTGTCGGGTCTTCTGTCGGGTCTCATGGGGTGTACGTGT 600
Db 580 CAATGAATGAATACCCAGACTCTCTGTCAGTACTGCTCAGACGTCATGGAGTATATGTGT 639
QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTGTATGACTCTCTTTGACA 660
Db 640 GGGGGGAAACATGGGAGAAAGGCCAAACCATGTGTGAGTGTGTATGACTTATTTTGA 699
QY 661 TTGCTCTCTCCATGAGAGAGATGGGACTCGATCCACACAGCTCCCGAGTTGGAGAAAATG 720
Db 700 TTGCCGTATCAATGAGAAAGTAGGACTTGTATCTTCACAGCTCCCGAGTTGGAGAAAATG 759
QY 721 GAATGTGTGAAGCAAGTGGATGCTTAAAGCATCTCCACAAATAAAAC-AAATCAATTAT 779
Db 760 GAATGTGTGAAGCAAGAAAGAGTCTAATATATATACAGAGATAAAGCTAAACGTAATTAT 819
QY 780 GCCTTAATAAATCACTCAGCTGCTTTTAAAAAATAAA 816
Db 820 TATTTAAATGAAAGCTATTTTAAAAATGAATGAAA 856
|||||

RESULT 7

BC017594

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

BC017594 1135 bp mRNA linear PRI 03-OCT-2003
Homo sapiens likely ortholog of mouse monocyte macrophage 19, mRNA
(cdna clone MGC:714 IMAGE:298833), complete cds.

BC017594 2 GI:34782764
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1135)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Utskin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abrams,R.D., Mullah,J.S., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 1135)

Strausberg,R.

TITLE Direct Submission
JOURNAL Submitted (21-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT On Sep 16, 2003 this sequence version replaced gi:17068426.
 Contact: MGC help desk
 Email: gcapsbe-x@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadams@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 3 Row: J Column: 19
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7705723.
FEATURES
 source
 1..1135
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:714 IMAGE:2988353"
 /tissue_type="Colon, adenocarcinoma"
 /clone_lib="NIH MGC_15"
 /lab_host="DH10B-R"
 /notes="Vector: pOTB7"
 1..1135
 /gene="MMRP19"
 /db_xref="LocusID:51074"
 37..765
 /codon_start=1
 /product="CGI-29 protein"
 /protein_id="AAH17594.1"
 /db_xref="GI:17068427"
 /db_xref="LocusID:51074"
 /translation="MSGCDWEGDCSRRCGAQDKHPYLIPELCKQPHLQWVLTGGISLKHGDEIYIAGSGVOKRQEDMFVINDENKDISGPSKRLKXSCQPLPFM NAYTMGAGAVIHTHKAAMATLFPREFKITHQEMIKKIKKSGGYRDDMLV VPIENTPEKDLKDRMAHMYPPSCAVLVRHGVYVWGETWEKAKTMCCEYDLF DIAVSMKKVGLDPSQLPVGENGIV"
 112..741
 /note="Aldolase II; Region: Class II Aldolase and Adducin N-terminal domain. This family includes class II aldolases and adducins which have not been ascribed any enzymatic function"
 /db_xref="CDD:pfam00596"

ORIGIN
 Query Match 70.9%; Score 595.4; DB 9; Length 1135;
 Best Local Similarity 84.7%; Pred. No. 3.4e-137;
 Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;
 4 CCATGCTGGCTGTCAGCT---CAAGAGACTGTGCTCGCGCCGCTGTGGCGCGCAGG 60
 |||||
 35 CCATGCTGGCTGTCAGCTGTGGAGGAGACTGTGTTCCCGAGATGCGCGCGCAGG 94
 |||||
 61 ACAAGGAGCACCCTGATTCCTGATCCAGAACTTTTGCAACAGTTTACCATCTGGGCT 120
 |||||
 95 ACAAGGAGCATCCAAGATACCTGATCCAGAACTTTTGCAACAGTTTACCATTAGGCT 154
 |||||
 121 GGGTCACTGGCACTGGAGGGGGAATCAGTTGAAGCATGGAATGAATCTACATTGCTC 180
 |||||
 155 GGGTCACTGGCACTGGAGGAGGAATTAGTTGAAGCATGCGATGAATCTACATTGCTC 214
 |||||
 181 CCTCAGCGTGCAGGAGCGCATTCAGCAGAGACATGTTGTGTGATTAATG 240
 |||||

215 CTTCAGGAGTGCAAAAGGAACGAATTCAGCCTGAAGACATGTTGTTTATGATATAAATG 274
 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAAGAGCTGAAAAAAGCCAGCTGCACTCCTC 300
 |||||
 275 AAAAGGACATAAGTGGACCTTCGCCATCGAAGAAGCTAAAAAAGCCAGTGTACTCTC 334
 |||||
 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGAGTGATTCATACCCTCTAAAG 360
 |||||
 335 TTTTCATGAATGCTTATACCAATGAGAGGAGCAGTGCAGTGATTCATACCCTCTAAAG 394
 |||||
 361 CTGCTGTGATGGTACCTCTCTCTTCCAGGACAGAGGTTTAAAAATTACACATCAAGAGA 420
 |||||
 395 CTGCTGTGATGGCACCTCTCTCTTCCAGGACGGGAGTTTAAAAATTACACATCAAGAGA 454
 |||||
 421 TGATCAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATAGCATGATATGTTAG 480
 |||||
 455 TGATAAAGGAATAGGAATGTACTTCCGAGGGTATTATACATATGATGATGTTAG 514
 |||||
 481 TGTACTCTATTATGAGACACCTCTCTGAAGAGAAAGGATCTCAAGAAAGAGTGGCTCATG 540
 |||||
 515 TGGTACCCATTATTGAGAAATACACCTTGAGGAGAAAGACCTCAAAAGATAGAAATGGCTCATG 574
 |||||
 541 CCATGAATGAGTACCCAGACTCTGTGCGGCTTCTGTCCGGCTCATGCGGCTGTACGTGT 600
 |||||
 575 CAATGAATGATACCCAGACTCTGTGCGAGTACTGTGTCAGACGTCATGGAGTATATGTTGT 634
 |||||
 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTATGACTACCTGTTTGACA 660
 |||||
 635 GGGGGGAAACATGGGAGAAAGCCAAACCATGTGTGAGTGTATTATGACTATTTATTTGATA 694
 |||||
 661 TTGCTCTCTCCATGAAGAGATGGGACTCGATCCACACAGCTCCAGTTGGGAAATG 720
 |||||
 695 TTGCCCTATCAATGAAGAAAGTAGGACTTGTATCTTTCACAGCTCCAGCTTGGAGAAATG 754
 |||||
 721 GAATTGTGTAAGCCAAAGTGGATGCTTAAGCATCTCCAAACAATAAATCAATCAATTAT 779
 |||||
 755 GAATTGTGTAAGCCAAAGAAAGAGTCTTAATATATACAGATTAAGCTAAACGTAATTAT 814
 |||||
 780 GCCTTAATAAATACTCAGCTGCTGCTTTTAAAAAAGAAAAA 816
 |||||
 815 TATTTAAATGAAGCTATTTTAAATGAATTGAAA 851
 |||||

RESULT 8
 BD168027
 LOCUS
 DEFINITION Novel proteins, gene encoding the same and method of utilization thereof.
 ACCESSION BD168027
 VERSION BD168027.1 GI:27873839
 KEYWORDS WO 0226978-A/2.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1136)
 Sha,S., Mukai,H., Aoki,Y. and Nishi,Y.
 Novel proteins, gene encoding the same and method of utilization
 Patent: WO 0226978-A 2 04-APR-2002;
 JAPAN TOBACCO INC,SHIKEN SHA,HIDEHITO MUKAI,YOSHIKO AOKI,YOSHISUKE NISHI
 OS Homo sapiens (human)
 PN WO 0226978-A/2
 FD 04-APR-2002
 PF 27-SEP-2001 WO 2001JP008446
 PR 27-SEP-2000 JP 00P 294191
 PI SHIKEN SHA,HIDEHITO MUKAI,YOSHIKO AOKI,YOSHISUKE NISHI PC
 C12N15/12, C12N1/21, C07K14/705, C07K16/28, A61K38/00, A61K39/395, PC
 A61K45/00,
 G01N33/566//
 A61K48/00, A61P7/00, A61P31/00, A61P43/00, G01N33/15, G01N33/50, PC
 PC C12P21/08, (C12N15/12, C12R1:91), (C12N1/21, C12R1:19) CC Human
 normal brain tissue

```

FH Key Location/Qualifiers
FT source 1..1136
FT /organism='Homo sapiens (human)'

FEATURES
    source
        Location/Qualifiers
        1..1136
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"

ORIGIN
    Query Match 70.9%; Score 595.4; DB 6; Length 1136;
    Best Local Similarity 84.7%; Pred. No. 3.4e-137;
    Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGCTCAAGCT---CAAGGAGACTGTGTGCTCGGCGCGTGTGGCGGCAGG 60
DB 44 CCATGCTGGCTGTGATGCTTGGAGGGAGACTGTGTGCTCCGAGATGCGGCGCGAGG 103
QY 61 ACAAGGAGCACCCCGATTCCTGATGCCAGAACTTTGCGAAACAGTTTACCATCTGGCT 120
DB 104 ACAAGGAGCATCCAGATACCTGATCCAGAACTTTGCGAAACAGTTTACCATTTAGGCT 163
QY 121 GGTCTACTGCACTGAGGGGGAATCAGCTTGAAGCATGCGATGAATCTACATTCCTC 180
DB 164 GGTCTACTGGAATGAGGAGGAATAGCTTTGAAGCATGCGATGAATCTACATTCCTC 223
QY 181 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGATGATTAATG 240
DB 224 CTTCAAGAGTGCAAAAGGAGCAATTCAGCTTGAAGCATGCGATGAATCTACATTCCTC 283
QY 241 AGCAGGACATAGCGGGCTTCAGCATCTTGAAGCTGAAAAGCGAGTGCATCTCCTC 300
DB 284 AAAAGGACATAGTGAGCTTCCCATCGAAGAGCTTAAAAAGCCAGTGTACTCCTC 343
QY 301 TTTTCTATGATGCTTATACATGAGAGGAGCTGCGCAGTGTATACCATCTCAAAG 360
DB 344 TTTTCTATGATGCTTATACATGAGAGGAGCTGCGCAGTGTATACCATCTCAAAG 403
QY 361 CTGCTGTGATGCTACCTTCTGTTTCCAGGACAGGATTTAAATTTACATCAAGAGA 420
DB 404 CTGCTGTGATGCTACCTTCTGTTTCCAGGACAGGATTTAAATTTACATCAAGAGA 463
QY 421 TGATCAAGAGATTAAGAAATGATACCTCAGGAGGCTATTACAGATAGCATGATGTTAG 480
DB 464 TGAATAAAGGATTAAGAAATGATACCTCAGGAGGCTATTACAGATAGCATGATGTTAG 523
QY 481 TGCTACTATTATTGAGAACACTCTGAGAGAGGATCTCAAGAGAAAGGATGGCTCATG 540
DB 524 TGTATCCCATTTTGAAGATACACCTGAGGAGAAAGACCTCAAGATAGATGGCTCATG 583
QY 541 CCATGAATGATGATACCCAGACTCCTGTGCGGTTCTTGTCCGCGCTCATGGGGTACGTT 600
DB 584 CAAATGAATGAATACCCAGACTCCTGTGCGAGTCTGTCAGACGTCATGGAGTATATGTT 643
QY 601 GGGGAGAACATGGGAGAAAGAAACCATGTTGAGTGTATGACTACCTGTTTGACA 660
DB 644 GGGGGGAAACATGGGAGAGGCCAAACCATGTTGAGTGTATGACTATTATTATTGATA 703
QY 661 TTGCTGTCTCCATGAAGACATGGGCTCGATCCACACAGCTCCCATGTTGGAGAAAATG 720
DB 704 TTGCTGTCTCCATGAAGACATGGGCTCGATCCACACAGCTCCCATGTTGGAGAAAATG 763
QY 721 GAATTTGTGAAGCAAGTGAATGCTTAAGCATCTCCAAACAATAAAG - AAATCTCAATTAT 779
DB 764 GAATTTGTGAAGCAAGTGAATGCTTAATTTATATACAGATTAAGCTAAACGTAATTAT 823
QY 780 GCCTTAATAAATCTAGCTGCTTTTAAAAAATAA 816
DB 824 TATTTAAATGAAGCTATTTTAAAAATGAATTGAAA 860

RESULT 9
E28764/c
```

Db 575 TGATAAAGGAATAAAGAAATGTACTTCCGGAGGTAATATAGATATGATGATGTTAG 516

Qy 481 TGGTACCTATTATTGAGAACACCTCTGAAGAGAGAGGATCTCAAGAAAGGATGCTCATG 540

Db 515 TGGTACCCATTATTGAGAAATACACCTGAGGAGAAAGACCTCAAGATAGAAATGCTCATG 456

Qy 541 CCATGAATGATGATCCAGACCTCCGTGCGGTTCTTGTCCGGGCTCATGGGTTGATGCTGT 600

Db 455 CAATGAATGATGATCCAGACCTCCGTGCGGTTCTTGTCCGGGCTCATGGGTTGATGCTGT 396

Qy 601 GGGGAGAGAAACATGGGAGAGAAACCAACCTGTGTGAGTGTATGACTTACTGTTTGCACA 660

Db 395 GGGGGGAGAAACATGGGAGAGAGGAGCAACCACTGTGTGAGTGTATGACTTATTGATTA 336

Qy 661 TTGCTGTCTCCATGAAGAAGATGGGATCGCATCGATCAACACAGCTCCCGAGTTGGAGAAATG 720

Db 335 TTGCCGTATCAATGAAGAAGTAGGACTTGTATCTTCCAGCTCCCGAGTTGGAGAAATG 276

Qy 721 GAATTTGTGAGCAAGTGGATGCTTGAAGCATCTCCAGCATTAACAACTCAATATAT 779

Db 275 GAATTTGTGAGCAAGTGGATGCTTGAAGCATCTCCAGCATTAACAACTCAATATAT 216

Qy 780 GCCTTAATAAATCACTGAGTGTCTTTTAAAAA 816

Db 215 TATTAAATGAAGCTATTTTAAATGAATGAA 179

BC008440 1199 bp mRNA linear PRI 07-OCT-2003

RESULT 10

LOCUS BC008440

DEFINITION Homo sapiens likely ortholog of mouse monocyte macrophage 19, mRNA

ACCESSION BC008440

VERSION BC008440.1 GI:14250070

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1199)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Vialalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 1199)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgapsb-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) med@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAL Plate: 21 Row: B Column: 5.

Location/Qualifiers

1. 1199

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:14646 IMAGE:4096399"

/tissue_type="Bone marrow, chronic myelogenous leukemia"

/clone_lib="NIH MGC_54"

/lab_host="DH10B"

/note="Vector: pDNR-LIB"

1. 1199

/gene="MRP19"

/db_xref="LocusID:51074"

95. 823

/codon_start=1

/product="CGI-29 protein"

/protein_id="AA08440.1"

/db_xref="GI:14250071"

/db_xref="LocusID:51074"

/translation="MSGCDAREGDCSSRRCAQDKERPRYLIPELCKQFYHLGVTGTGGGSLKHGDEIVAPSGVOKERIQEDMFVDINEKDISGSPSKKXKSLQCTPLPMNAYTMRGAGAVIHTSKAAVWATLPPREFKITHQEMIKIKTKTSGGYRVDMLVVPILNTPPEKDLKDRMAHAYNEPDSAVLVRRHGYVWGTWEKAKTWCECYDILFDIAVSKKVGLDPSQUPVGNGV"

170. 799

/note="Aldolase II; Region: Class II Aldolase and Adducin N-terminal domain. This family includes class II aldolases and adducins which have not been ascribed any enzymatic function"

/db_xref="CDD:pfam00596"

Query Match 70.7% Score 593.8; DB 9; Length 1199;

Best Local Similarity 84.6% Pred. No. 8.4e-137;

Matches 691; Conservative 0; Mismatches 122; Indels 4; Gaps 2;

Qy 4 CCATGCTGCTGTCAAGCTC--AAGGAGACTGTTGCTCGGGCCGTGTGGCGCGCAGG 60

Db 93 CCATGCTGCTGTGATGCTCGGGAGGAGACTGTTGTTCCCGGAGATGCGCGCGCAGG 152

Qy 61 ACAAGAGACACCCCGGATTCCTGATCCAGAACTTTGCCAAACAGTTTTTACCATCTGCGCT 120

Db 153 ACAAGAGCGTCCAGATACCTGATCCAGAACTTTGCCAAACAGTTTTTACCATTAGGCT 212

Qy 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTTGAAGCATGGCAATCTACATTCCTC 180

Db 213 GGGTCACTGGCACTGGAGGAGGAATAGCTTTGAAGCATGGCAATCTACATTCCTC 272

Qy 181 CTTGAGCGCTGCAAAAGGAGCGCATTCAGCCAGACATGTTTGTGTGTGACATTATG 240

Db 273 CTTGAGCGCTGCAAAAGGAGCGCATTCAGCCAGACATGTTTGTGTGTGACATTATG 332

Qy 241 AGCAGGACATAGCGGCTCCAGCATCTTAAGAGCTTGAAGCTGAAAAAGCAGCTGCTC 300

Db 333 AAAAGGACATAGTGGACCTTCGCCATCGAAGAGCTTAAAAAGCAGCTGACTCTC 392

Qy 301 TTTTTCATGAATGCTTATACCATGAGAGGAGCTCGCGCAGTGTATCATACCCTCTAAG 360

Db 393 TTTTTCATGAATGCTTATACCATGAGAGGAGCTCGCGCAGTGTATCATACCCTCTAAG 452

FEATURES

source

gene

CDS

misc_feature

ORIGIN

361	QY	CTGCTGTGATGGCTTACCGCTTCTGTTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA	420
453	Db	CTGCTGTGATGGCCACACTTCTCTTTTCCAGACGGGAGTTTAAATTTACATCAAGAGA	512
421	QY	TGATCAAAGGAATAAGCAAAATGTACCTCAGGAGCTATTACAGATACGATGATGTTAG	480
513	Db	TGATAAAGGATAAAGAAATGATTTCTCCGGAGGTTATTAGATATGATGATGTTAG	572
481	QY	TGGTACCTATTATTGTGAAACACACTCCTGAAGAGAAAGGATCTCAAGAAGAGATGGCTCATG	540
573	Db	TGGTACCCATTATTGTAGAATAACACTGAGGAGAAAGACCTCAAAAGATAGAATGGCTCATG	632
541	QY	CCATGAATGAGTACCCAGACTCCGTGCGGTTCTTGTCGCGCGTCATGGGGTGTAACGTGT	600
633	Db	CAGTGAATTAATACCCAGACTCCTCTGTGCGAGTACTGGTCAGACGTCATGGAGTATATGTGT	692
601	QY	GGGAGAGAAACATGGGAGAAAGCAAAAACCAATGTGTGAGTGTATTGACTACTCTGTTGACA	660
693	Db	GGGGGAAAACATGGGAGAGGCCAAAACCAATGCTGTGAGTGTATTGACTATTATTTTGATA	752
661	QY	TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAACACAGCTCCCGATGGGAGAAATG	720
753	Db	TTGCGGTATCAATGAAGAAAGTAGGACTTGATCTTCCAGCTCCCGATGGGAGAAATG	812
721	QY	GAATTGTGAAGCCCAAGTGGATGCTTAAGCATCTCCAACAATAAACC-AAACTCAATTTAT	779
813	Db	GAATTGTCTAAGCCAAAGAAAGTCTTAATTATATACAGATTAAGCTTAACGTAAATAT	872
780	QY	GCCTTAATAAATACTAGCTGCTTTTAAAAAATAA	816
873	Db	TATTTAATGAAGACTATTTTTTAAATGAATGAAA	909

RESULT_11				
BD107338	BD107338	1226 bp	DNA	linear
LOCUS				
DEFINITION	Novel chondrogenesis promoter.			
ACCESSION	BD107338			
VERSION	BD107338.1	GI:23202156		
KEYWORDS	JP 2002020311-A/1.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 1226)
TITLE	Kato,Y. and Fujimoto,K.
JOURNAL	Novel chondrogenesis promoter
COMMENT	Patent: JP 2002020311-A 1 23-JAN-2002; JAPAN SCIENCE AND TECHNOLOGY CORP OS Homo sapiens (human) PN JP 2002020311-A/1 PD 23-JAN-2002 PF 07-JUL-2000 JP 2000206566 PI YUKIO KATO, KATSUMI FUJIMOTO PC A61K38/00,A61P19/00,A61P19/02,A61P19/08,C07K16/18,C12N15/09, C12Q1/68, PC G01N33/53,G01N33/566//C07K14/42,C07K14/51,C07K14/79,A61K37/02, C12N15/00 CC Novel chondrogenesis promoter FH Key Location/Qualifiers FT source 1..1226 FT /organism='Homo sapiens (human)'. Location/Qualifiers 1..1226 FEATURES SOURCE /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"

Query Match	70.7%	Score 593.8	DB 6	Length 1226
Best Local Similarity	84.6%	Pred. NO. 8.4e-137		
Matches 691	Conservative	0	Mismatches 122	Indels 4
				Gaps 2

QY	4	CCATGCTCGCTGTCAAGCT---CAAGGACACTCTTGCTCGCGCGCTGTGGCGCGCAGG	60
Db	76	CCATGCTCGCTGTGATGCTGGGAGGAGGACACTGTGTTCCCGGAGATGCGGCGCGCAGG	135
QY	61	ACAAGGAGCACCCCGGATTCCTGATCCCAAGAACTTTTGCACAAACAGTTTACCATCTGGGCT	120
Db	136	ACAAGGAGCATCCAAGATACCTGATCCCGAACTTTGCAAAACAGTTTACCATTTAGGCT	195
QY	121	GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGSCATGAATCTACATTCGCTC	180
Db	196	GGGTCACTGGGACTGGAGGAGGAATAGCTTGAAGCATGGCGATGAATACTACATTCGCTC	255
QY	181	CCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGAGAATGTTTGTGTGTGACATTAATG	240
Db	256	CTTCAGGAGTCAAAAGGAACGAATTCAGCCTGAGACATGTTGTTGTGTGATATAATG	315
QY	241	AGCAGGACATGAAGCGGCGCTCCAGACTCTAAGAAAGCTGAAAGAACGAGTGCATCTCCTC	300
Db	316	AAAAGGACATAAGTGGAGCTTTCGCATCTGGAAGAGCTTAAAAAAGACGAGTGTACTCCTC	375
QY	301	TTTTTCATGAATGCTTATACCATGAGAGGAGCTGCGCAGTGATTTCATACCCACTCTAAAG	360
Db	376	TTTTTCATGAATGCTTACCAATGAGAGAGCAGGTGCGAGTATTCATCCCATCTTAAG	435
QY	361	CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAAAATTCACATCAAGAGA	420
Db	436	CTGCTGTGATGGCCACACCTTCTCTTCCAGGACGGGAGTTTAAAAATTCACATCAAGAGA	495
QY	421	TGATCAAGAGATTAAGAAATGTACCTCAGAGGCTATTACAGATACGATGATATGTTAG	480
Db	496	TGATAAAGAGATTAAGAAATGTACTTCCGGAGGGTATTATAGATATGATGATGTTAG	555
QY	481	TGGTACCTATTATTAGAAACACTCCTCAAGAGAGAGATCTCAAGAAAGGATGGCTCATG	540
Db	556	TGGTACCCATTATTGAGAAATACACTCAGGAGAGACCCCTCAAGATAGAATGGCTCATG	615
QY	541	CCATGAATGAGTACCCAGACTCTCTGTGCGGGTCTTGTTCGGCGCTCATGGGGTGTACGTGT	600
Db	616	CAATGAATGAATACCCAGACTCTCTGTGCTGACTGCTGAGACGCTCAGGAGTATATGTGT	675
QY	601	GGGCGAGAAACATGGGAGAGAAACCACTGTGTGAGTGTATTGACTACTGTTTGCACA	660
Db	676	GGGGGAGAAACATGGGAGAGGCGAAACCAATGTGTGAGTGTATTGACTATTTATTGATA	735
QY	661	TTGCTGTCTCATGAAGAAGATGGGACTCGACTCCACACAGCTCCCGAGTGGAGAAAATG	720
Db	736	TTGCGGTATCAATGAAGAAAGTAGGACTTGCATCTTCACAGCTCCCGAGTGGAGAAAATG	795
QY	721	GAATGTGTAGCCAGTGGATGCTGAAGCATCTCCAACAATAAAGC-AAAATCAATAT	779
Db	796	GAATGTGTAGCCAAAGAAAGTCTAATTAATACAGATTAAGCTAAACGTAATAT	855
QY	780	GCCTTAATAAATCAGCTGCTTTTAAAAAATAAATAA 816	
Db	856	TATTTAAATGAAGCTATTTTTTAAATGAATGAAA 892	
RESULT 12			
AF132963			
LOCUS	AF132963	1226 bp	linear
DEFINITION	Hom sapiens CGI-29 protein mRNA, complete cds.		PRI 18-MAY-2000
ACCESSION	AF132963		
VERSION	AF132963.1		
KEYWORDS	GI:4680696		
SOURCE	Hom sapiens (human)		
ORGANISM	Hom sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1226)		
JOURNAL	Lai, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S. and Lin, W.		
	Identification of novel human genes evolutionarily conserved in		
	Cenorhabditis elegans by comparative proteomics		
	Genome Res. 10 (5): 703-713 (2000)		

MEDLINE 20272150
PUBMED 10810093
REFERENCE 2 (bases 1 to 1226)
AUTHORS Lin, W.-C.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1999) Institute of Biomedical Sciences, Academia Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan
FEATURES
source Location/Qualifiers
1..1226
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
78..806
/codon_start=1
/product="CGI-29 protein"
/protein_id="AAD27738.1"
/db_xref="GI:4680697"
/translation="MSGCDAGEBDCSRRCGQAQKHEPRYLILPELCKQFVHLGWTGTGGISLKHGAVIYTHSKAAVATLPPGRFKITHQEMIKIKKSGGYRVDMLVVPILNTPEKTLKDRMAHANEVDPSCAVLVRHGVYVWGETWEKAKTWCECYDLFDIAVSMKKVGLDPSQLPVGNGIV"
ORIGIN
Query Match 70.7%; Score 593.8; DB 9; Length 1226;
Best Local Similarity 84.6%; Pred. No. 8.4e-137;
Matches 691; Conservative 0; Mismatches 122; Indels 4; Gaps 2;
QY 4 CCATGTCGTGTCAGCT---CAAGAGACTGTTGCTCGCGCGCTGTGGCGCGCAGG 60
Db 76 CCATGTCGTGTCAGCTGCGGAGGAGACTGTTGTTCCCGAGATGCGCGCGCAGG 135
QY 61 ACAGGAGACCCCGGATCTCTGATCCAGAACTTCCAAACAGTTTACCATCTGGCT 120
Db 136 ACAGGAGACCTCCAAAGATACCTGATCCAGAACTTCCAAACAGTTTACCATCTAGCT 195
QY 121 GGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATCAAATCTACATTGCTC 180
Db 196 GGGTCACTGGCACTGGAGGAGTAAGCTTGAAGCATGGCAATCAAATCTACATTGCTC 255
QY 181 CCTCAGCGGTGCAAAAGAGCGCAATTCAGCAGAGACATGTTTGTGTGACATTAATG 240
Db 256 CTTCAGGAGTGCAAAAGCAAGATTCAGCCTGAAGACATGTTTGTGTGATATAATG 315
QY 241 AGCAGGACATAGCGGCGCTCCAGCATCTAGAGAGCTGAAAAGAGCAGTGCACTCTC 300
Db 316 AAAAGGACATAGTGGACCTTCGCGCATCGAAGAGCTTAAAAAGCCAGTGTACTCTC 375
QY 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGGCGAGTCAATCATACCCACTCTAAAG 360
Db 376 TTTTCATGAATGCTTACCAATGAGAGAGCAGGTGCGAGTCAATCACTCAAG 435
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCAGACAGGAGTTTAAATTCACATCAAGAGA 420
Db 436 CTGCTGTGATGGCCACACTTCTCTTCCAGACGGGAGTTTAAATTCACATCAAGAGA 495
QY 421 TGATCAAGGAATGAAGAAATGTAACCTCAGGAGCTATTACAGATACATCATATGTAG 480
Db 496 TGATAAAGGAATGAAGAAATGTAATTCAGGAGGTATTAAGATATGATATGTAG 555
QY 481 TGGTACTTATTATTGAGAACTCTCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
Db 556 TGGTACCATTATTGAGAACTACCTGAGGAGAGACCCCTCAAGATAGATGGCTCATG 615
QY 541 CCATGAATGATACCCAGACTCTGTGGGTCTTGTTCGGGGTCAATGGGGTGTACGTGT 600
Db 616 CAATGAATGAATACCCAGACTCTGTGGGTCTTGTTCGGGGTCAATGGGGTGTACGTGT 675
QY 601 GGGGAGAACTGGGAGAGACCAAAACCATGTGTGAGTGTATGACTACCTGTTGACA 660
Db 676 GGGGGGAAACATGGGAGAGGCCAAACCATGTGTGAGTGTATGACTATTATTGTATA 735
661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTGGAGAAATG 720

Db 736 TTGCCGTATCAATGAAGAAAGTAGGACTTGATCTTCACAGCTCCAGTGGAGAAATG 795
QY 721 GAATTTGTGAAGCAAGTGCATGCCAAGCATCTCCAACTAAAC-AAAATCAATTAT 779
Db 796 GAATTTGTGAAGCAAGAAAGTCTAAATTTATACAGATTAAGCTAAACGTAATTAT 855
QY 780 GCCTTAATAAAACTCAGCTGCTTTTAAAAAAGAAAA 816
Db 856 TATTAAATGAAGCTATTTTAAATGAATTGAAA 892
RESULT 13
AL807761 103861 bp DNA linear PRI 08-JAN-2003
LOCUS Human DNA sequence from clone RP11-196118 on chromosome 9, complete
DEFINITION sequence.
ACCESSION AL807761
VERSION AL807761.8 GI:27645838
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Lloyd, D.
Direct Submission
Submitted (08-JAN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Jan 11, 2003 this sequence version replaced gi:22416103.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-196118 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VBCOR: pBACe3.6
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
FEATURES
source Location/Qualifiers
1..103861
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-196118"
/clone_lib="RP11-11.1"

ORIGIN


```
Query Match      68.6%; Score 576.2; DB 9; Length 103861;
Best Local Similarity 83.2%; Pred. No. 3.1e-132;
Matches 680; Conservative 0; Mismatches 133; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGTCAAGTCTC---AAGGAGACTGTGTGCTCGCGGCGGTGTGGCGGCGAGG 60
Db 91885 CCATGCTGGCTGTGAGCGCCGGGAGGAGACTGTGTGTTCCCGAGATGCGGAGCGCAGG 91944

QY 61 ACAAGGAGACACCCCGATTCCTGATGCCAGAACTTTGCAACAGTTTACCATCTGGGCT 120
Db 91945 ACAAGGAGATCCCAAGATACCTGATGCCAGAACTTTGCAACAGTTTACCATCTGGGCT 92004

QY 121 GGGTCACTGGCACTGAGGGGGAATCAGCTTCAAGCATGGCAATCAATCTCAATCTGCTC 180
Db 92005 GGGTCACTGGCACTGAGGAGGAATGCTTGAAGCATGGCGGTGAATCTCAATCTGCTC 92064

QY 181 CCTCAGCGGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTTGTGTGTCATTAATG 240
Db 92065 CTTCAGGAGTGCAAAAGGAACGAATTCAGCTCTAAGAGCTAAATTAAGCCAGTGTACTCTC 92124

QY 241 AGCAGACATAGCGGGCTCCAGCATCTAAGAGCTGAAGAAAGCCAGTCACCTCCCTC 300
Db 92125 AAAGGACATAGTGAGACCTTCGCCATCTAAGAGCTAAATTAAGCCAGTGTACTCTC 92184

QY 301 TTTTTCATGAATGCTTATACCATGAGAGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 360
Db 92185 TTCTCATGATGCTTACCAATGAGAGAGCAGGTGCGGTGATTCATACCCACTCTAAAG 92244

QY 361 CTGCTGTGATGCTACCTCTTCTGTTTCAGGACAGAGTGTAAATTAACATCAAGAGA 420
Db 92245 CTGCTGTGATGCTACCTCTTCTGTTTCAGGAGTGTAAATTAACATCAAGAGA 92304

QY 421 TCATCAAGCAATAGGAATCTACTCAGAGGCTTATACAGATACGATGATGTTAG 480
Db 92305 TGATCAAGGAATAAGAAATGTACTCTGGAGGGTATTATAGATGATGATGTTAG 92364

QY 481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAAGATCTCAAGAAAGGATGGCTCATG 540
Db 92365 TGGCACCCATTATTGAGATACACCTGAGGAGAGAGATCTCAAGATGAGATGGCTCATG 92424

QY 541 CCATGAATGAGTACCAGACTCCTGTGCGGTTCTTGTCCGGCGTCAATGGGGTGTACGTG 600
Db 92425 CAATGAATGAATACCAGACTCCTGTGCGAGTACTGTCAGAGCTCATGAGTATATGTGT 92484

QY 601 GGGGAGAACATGGGAGAGCAAAACCATGCTGTGAGTGTATGACTACCTGTTTGACA 660
Db 92485 GGGGGAACATGAGAGAGGCGCAAAACCATGCTGTGAGTGTATGACTATTTTGTGATA 92544

QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCACACAGCTCCCGTTCGAGAAATG 720
Db 92545 TTGCGGTATCAATGAAGAAAGTAGACTTGTATCTTCAGCTCCCGTTCGAGAAATG 92604

QY 721 GAATGTGTAAGCAAGTGGATGCTTACGATCTCCACATATAA-ACAACTCAATAT 779
Db 92605 GAATGTGTAAGCAAGTGGATGCTTATATATATACAGAGATAAAGTTAAACATATAT 92664

QY 780 GCCTTAAATAAACTCAGCTGCTTTTAAAAAATAA 816
Db 92665 TATTAAATGAAGCTATTTTAAAAATGAATGAAA 92701
```

```
RESULT 14
AC068188 147844 bp DNA linear HTG 28-MAY-2000
LOCUS Homo sapiens chromosome 8 clone RP11-196118 map 8, WORKING DRAFT
DEFINITION SEQUENCE, 20 unordered pieces.
ACCESSION AC068188
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (Bases 1 to 147844)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-196118
Unpublished
2 (Bases 1 to 147844)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galligan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Labrecque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meidram, J., Meneus, D., Mihova, T., Miranda, C., Miengo, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 28, 2000 this sequence version replaced gi:7670231.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTHR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9711

Center clone name: 196_I_18

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 137250 bases at least Q40

Consensus quality: 142296 bases at least Q30

Consensus quality: 144570 bases at least Q20

Insert size: 157000; agarose-fp

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 20 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 6: contig of 6 bp in length

* 7 106: gap of 100 bp

* 107 1338: contig of 1232 bp in length

* 1339 1438: gap of 100 bp

* 1439 3068: contig of 1630 bp in length

* 3169 4802: contig of 1634 bp in length

* 4803 4902: gap of 100 bp

* 4903 8244: contig of 3342 bp in length

* 8245 8345: gap of 100 bp

* 11578: contig of 3234 bp in length

```
* 11579 11678: gap of 100 bp
* 11679 1391: contig of 2213 bp in length
* 1392 1391: gap of 100 bp
* 1392 1558: contig of 1567 bp in length
* 1559 1558: gap of 100 bp
* 1559 1958: contig of 4000 bp in length
* 1958 1958: gap of 100 bp
* 1959 23937: contig of 4179 bp in length
* 23937 24037: gap of 100 bp
* 24037 28435: contig of 4398 bp in length
* 28435 28436: gap of 100 bp
* 28436 32630: contig of 4095 bp in length
* 32630 32631: gap of 100 bp
* 32631 38632: contig of 5902 bp in length
* 38632 38633: gap of 100 bp
* 38633 43912: contig of 5180 bp in length
* 43912 43913: gap of 100 bp
* 43913 55699: contig of 11687 bp in length
* 55699 55700: gap of 100 bp
* 55700 70128: contig of 14328 bp in length
* 70128 70227: gap of 100 bp
* 70227 89897: contig of 19670 bp in length
* 89897 89997: gap of 100 bp
* 89997 111024: contig of 21027 bp in length
* 111024 111025: gap of 100 bp
* 111025 129265: contig of 18140 bp in length
* 129265 129365: gap of 100 bp
* 129365 147844: contig of 18480 bp in length.
FEATURES
    source
        1..147844
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="8"
            /map="8"
            /clone="RP11-196118"
            /clone_lib="RP11-196118 Human Male BAC"
        1..6
            /note="assembly_fragment"
            clone_end:T7
            vector_side:right
        107..1338
            /note="assembly_fragment"
            1439..3068
                /note="assembly_fragment"
            3169..4802
                /note="assembly_fragment"
            4903..8244
                /note="assembly_fragment"
            8345..11578
                /note="assembly_fragment"
            11679..13891
                /note="assembly_fragment"
            13992..15558
                /note="assembly_fragment"
            clone_end:SP6
            vector_side:right
        15659..19658
            /note="assembly_fragment"
        19759..23937
            /note="assembly_fragment"
        24038..28435
            /note="assembly_fragment"
        28536..32630
            /note="assembly_fragment"
        32731..38632
            /note="assembly_fragment"
        38733..43912
            /note="assembly_fragment"
        44013..55699
            /note="assembly_fragment"
        55800..70127
            /note="assembly_fragment"
```

```
misc_feature 70228..89897
    /note="assembly_fragment"
misc_feature 89998..111024
    /note="assembly_fragment"
misc_feature 111125..129264
    /note="assembly_fragment"
misc_feature 129365..147844
    /note="assembly_fragment"
ORIGIN
    Query Match 68.6%; Score 576.2; DB 2; Length 147844;
    Best Local Similarity 83.2%; Pred. No. 3.2e-132; Indels 4; Gaps 2;
    Matches 680; Conservative 0; Mismatches 133;
QY 4 CCATGTCTGGCTGTCAAGCTC---AAGGAGAGTGTGTCTCGCGGCGGTGTGGCGGCAGG 60
DB 127305 CCATGTCTGGCTGTGACGCCCGGGAGGAGACTGTGTGTTCCCGAGATCGCGAGCGCAGG 127365
QY 61 ACAAGGACACCCCGGATTCCTGATCCCGAAGTCTTGCAACAGTGTACCATCTGGGCT 120
DB 127365 ACAAGGACATCCAAGATACCTGATCCCGAAGTCTTGCAACAGTGTACCATCTGGGCT 127425
QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTTGCTC 180
DB 127425 GGGTCACTGGCACTGGAGGAGGAATTTGGCTTGAGGCATGGCGGTGAATCTACATTTGCTC 127485
QY 181 CCTCAGGGGTGCAAAAGGAGCGCATTCAGCCAGAAGACATGTTGTGTGTGACATTAATG 240
DB 127485 CTTTCAGGAGTGCAAAAGGAGCGAATTCAGCTGAAGACATGTTGTGTGTGATATAATG 127545
QY 241 AGCAGGACATAAGCGGGCTTCAGCATCTAAGAGCTGAAAAAAGCCAGTCGCTCCTC 300
DB 127545 AAAAGGACATAGTGGACCTTCGCCATCTAAGAGCTAAAAAAGCCAGTCGCTCCTC 127605
QY 301 TTTTCATGAATGCTTTATACCATGAGAGGAGCTGGCGCAGTGTTCATACCCACTCTAAAG 360
DB 127605 TTCTCATGAATGCTTACACATGAGAGGAGCAGTGGCGGTGATTCATACCCACTCTAAAG 127665
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTAAAAATTCACATCAAGAGA 420
DB 127665 CTGCTGTGATGGCCACCCTTCTCTTTTCGAGGATGGAGTTAAAAATTCACATCAAGAGA 127725
QY 421 TGATCAAGGAATAAGGAATGTACTCTCAGAGGCTATTACAGATACGATGATATGTTAG 480
DB 127725 TGATCAAGGAATAAGGAATGTACTCTCAGAGGCTATTATAGATATGATATGTTAG 127785
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAGAGAGAGTCTCAAGAGAGAGTGGCTCATG 540
DB 127785 TGGCACCATTTATTGAGATACACCTGAGGAGAGAGACTTCAAGATAGATGAGTCTCATG 127845
QY 541 CCATGAATGAGTACCAGACTCCTGTGGGGTTCTTGTCCGGCGTCATGGGGTGTAGTGT 600
DB 127845 CAATGAATGAATACCAGACTCCTGTGCAGTACTGTCAGAGCTCATGGAGTATATGCT 127905
QY 601 GGGGAGAAACATGGGAGAAAGCAAAAACCATGTGTGAGTGTATTGACTACCTGTTTGACA 660
DB 127905 GGGGGGAAACATGAGAGAGGCAAAACCATGTGTGAGTGTATTGACTATTTATTGATA 127965
QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACAGCTCCAGTTCGAGAAAAATG 720
DB 127965 TTGCGGTATCAATGAAGAGAGTGTGAGTGTATTGACTACCTGTTGAGAAAAATG 128025
QY 721 GAATTGTGTAAAGCCAAAGTGGATGCCCTTAAGCATCTCCAAACAATAAAA-ACAACTCAATTAT 779
DB 128025 GAATTGTCTAAGCCAAAGAGAGTCTAATTATATACAGAGATAAAGTTAAACATAATTAT 128085
QY 780 GCCTTAATAAATCTCAGCTGCTTTTAAAAAATAAAAA 816
DB 128085 TATTTAAATGAAGCTATTTTTTTAAATGAATTGAAA 128122
RESULT 15
AC068376/c
```


misc_feature	/note="assembly_fragment"																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	</
--------------	---------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

DB	96674	CTGCTGTGATGGCCACCCTTCTCTTTCGAGGATGGAGTTTAAAAATTACCCATCAAGAGA 96615
QY	421	TGATCAAAAGGAATAAGGAATGTACTCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
DB	96614	TGATCAAAAGGAATAAGGAATGTACTCTCTGGAGGCTATTATAGATATGATGATGTTAG 96555
QY	481	TGGTACCTATTATTGAGAACACTCTCTGAGAGAAAGATCTCAAGAAAGGATGGCTCATG 540
DB	96554	TGGCACCCATTATTGAGAATAACCTCTGAGGAGAAAGACTTCAAGATAGATGGCTCATG 96495
QY	541	CCATGAATGAGTACCCAGACTCTCTGTGGGTTCTTGTCCGGCTCATGCGGTGACGTCT 600
DB	96494	CAATGAATGATACCCAGACTCTCTGTGAGTACTGTCTAGACGTATGAGTATGTTGT 96435
QY	601	GGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACCTGTTTGACA 660
DB	96434	GGGGGAAACATGAGAGAAAGGCCAAACCATGTGTGAGTGTATTGACTATTATTGATA 96375
QY	661	TTGCTGTCTCCATGAGAAAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720
DB	96374	TTGCCGTATCAATGAAGAAAGTAGGACTTGTATCTTCCAGCTCCCGCTGGAGAAATG 96315
QY	721	GAAATTGTGTAAGCCAAAGTGGATGCCCTTAAGCATCTCCAAACAATAAA-ACAAACTCAATTAT 779
DB	96314	GAAATTGTCTAAGCCAAAGAAAGTCTTAATTATATACAGATATAAGTTAAACATTAATTAT 96255
QY	780	GCCTTAAATAAAACTCAGCTGCTTTTAAAAAATAAAAA 816
DB	96254	TATTTAAATGAAGCTATTTTTTTTAAATGAATTGAAA 96218

Search completed: June 2, 2004, 00:00:50
Job time : 5522 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 19:29:51 ; Search time 551 Seconds
(without alignments)
6476.382 Million cell updates/sec

Title: US-09-937-905-1

Perfect score: 840

Sequence: 1 gaaccatgctgctgtca.....aaaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	840	100.0	840	5	AAC61149
2	840	100.0	840	6	ABK12567 DNA encod
3	840	100.0	841	6	ABK47966 Murine cD
4	598.6	71.3	1183	5	Aah75155 Nucleotid
5	595.4	70.9	924	5	AAC61150 DNA encod
6	595.4	70.9	1136	6	ABK47967 Human cDN
7	595.4	70.9	1190	2	Aax05748 Nucleotid
8	593.8	70.7	1226	6	ABK12566 DNA encod
9	546	65.0	602	9	ADB51659 Primari r
10	399.8	47.6	3764	7	ABX63252 Human cDN
11	360	42.9	758	2	Aaz16856 Human gen
12	354	42.1	778	2	Aaz17165 Human gen
13	348.2	41.5	851	2	Aaz15675 Human gen
14	311.4	37.1	426	7	ABZ18972 Group III
15	256	30.5	796	4	ABL17337 Drosophil
16	248.8	29.6	300	2	AAZ12753 Human gen
17	246.4	29.3	479	8	ACH29056 Human adu
18	245.6	29.2	300	2	Aaz12752 Human gen
19	190.4	22.7	664	4	Aah33805 Human col
20	177.8	21.2	2934	4	ABU17336 Drosophil
21	177.8	21.2	10061	4	ABU17116 Drosophil
22	143	17.0	229	2	AAT08638 HPV E6-bi
23	143	17.0	229	2	AAT73916 E6-Bindin

ALIGNMENTS

RESULT 1

AAC61149
ID AAC61149 standard; DNA; 840 BP.

XX AC AAC61149;

XX AC

DT 07-FEB-2001 (first entry)

XX DE DNA encoding antigen recognised by Ab capable of inducing G-CSF activity.

XX DE Antigenic protein; antibody; granulocyte colony stimulating factor;

XX DE G-CSF; cancer therapy; bone marrow suppression; mouse; ds.

XX KW Mus sp.

XX OS WO2000060075-A1.

XX PN 12-OCT-2000.

XX PD 31-MAR-2000; 2000WO-JP002080.

XX PF 01-APR-1999; 99JP-00095092.

XX PR (NISB) JAPAN TOBACCO INC.

XX PA Sha S, Aoki Y, Nishi Y;

XX PI WPI; 2001-024452/03.

XX DR P-PSDB; AAY85635.

XX PT Gene encoding an antigen recognizing an antibody which induces

XX PT granulocyte colony stimulating factor (G-CSF) expression for gene therapy

XX PT and treatment of G-CSF associated disorders e.g. the side effects of

XX PT cancer therapy.

XX PS Claim 1; Page 47-49; 58pp; Japanese.

XX CC The present invention relates to a gene encoding an antigenic protein

XX CC recognised by an antibody or its fragments which can induce the

XX CC production of granulocyte colony stimulating factor (G-CSF). Also

XX CC included in the invention are partial sequences of the gene, antibodies

XX CC recognising all or part of the antigenic protein, expression vectors

XX CC containing the gene and host cells transformed by the vector. The gene is

XX CC used for gene therapy, and compounds identified by screening using the

XX CC gene sequence are used for treatment and prevention of disorders

XX CC associated with G-CSF expression such as the side effects of cancer

Aai90714 Human pol
Aac94027 Cat flea
Ab166584 Lung canc
Abn51202 Mouse spl
Abz08215 Human leu
Abz5566 Aspergill
Abk39958 Human che
Abv23003 Human imm
Abv23003 Human pro
Abv28839 Human pro
Abv55889 Human pro
Aax25130 Soybean i
Aaa38425 Soybean i
Aaf22334 Human sec
Aax52274 Protein P
Aac78688 Human PRO
Aaf72432 Human PRO
Aas45936 Human DNA
Abx78539 Human PRO
Aca59212 Human PRO
Aca75511 Novel hum
Aca70991 Human sec

QY 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTCTTAAAG 360
 Db 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTCTTAAAG 360
 QY 361 CTGCTGTGATGGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
 Db 361 CTGCTGTGATGGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
 QY 421 TGATCAAGGAATTAAGGAATGACCTCAGGAGGCTTATACAGATACGATGATGTAG 480
 Db 421 TGATCAAGGAATTAAGGAATGACCTCAGGAGGCTTATACAGATACGATGATGTAG 480
 QY 481 TGGTACCTATTATTGAGAACACTCCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
 Db 481 TGGTACCTATTATTGAGAACACTCCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
 QY 541 CCATGAATGAGTACCCAGACTCTGTCGGTCTTCTGCGGCTCATGGGCTGACGTGT 600
 Db 541 CCATGAATGAGTACCCAGACTCTGTCGGTCTTCTGCGGCTCATGGGCTGACGTGT 600
 QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
 Db 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
 QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720
 Db 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720
 QY 721 GAATTTGTAAAGCAAGTGTGCTTAAAGCAATCCAAACATCAAACTCAATATG 780
 Db 721 GAATTTGTAAAGCAAGTGTGCTTAAAGCAATCCAAACATCAAACTCAATATG 780
 QY 781 CCTTAATAAACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 840
 Db 781 CCTTAATAAACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 840

RESULT 3

ID ABK47966 standard; cDNA; 841 BP.
 XX ABK47966;
 AC
 XX
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Marine cDNA encoding G-CSF-inducible antibody binding protein, MMR19.
 XX
 XX Mouse; granulocyte-colony stimulating factor; G-CSF; ss; gene; MMR19;
 KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;
 KW infection.
 XX
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH 6..731
 FT CDS /tag= a
 FT /product= "G-CSF-inducible antibody binding protein"
 XX
 PN W0200226978-A1.
 XX
 XX
 PD 04-APR-2002.
 XX
 XX 27-SEP-2001; 2001WO-JP008446.
 XX
 XX 27-SEP-2000; 2000JP-00294191.
 XX
 XX (NISB) JAPAN TOBACCO INC.
 XX
 XX Sha S, Mukai H, Aoki Y, Nishi Y;
 PI WPI; 2002-340016/37.
 DR P-PSDB; AAU71177.

XX Gene encoding protein binding to antibody having granulocyte-colony
 PT stimulating factor (G-CSF) inducing activity, useful for screening
 PT potential drugs treating G-CSF associated diseases.
 XX
 PS Claim 2; Page 91-93; 103pp; Japanese.
 XX
 CC The invention relates to a mouse or human gene (MMR19) encoding a protein
 CC which binds to antibodies or their fragments which induce granulocyte-
 CC colony stimulating factor (G-CSF) secretion. The genes and proteins of
 CC the invention are used in diagnosis, treatment and prevention of diseases
 CC associated with G-CSF, including infections and neutrophil deficiency
 CC disease. This sequence represents a cDNA encoding a mouse G-CSF-inducible
 CC antibody binding protein, MMR19
 XX
 SQ Sequence 841 BP; 274 A; 174 C; 203 G; 190 T; 0 U; 0 Other;

Query Match 100.0%; Score 840; DB 6; Length 841;
 Best Local Similarity 100.0%; Pred. No. 2e-183;
 Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTGCTCGGGCGCTGTGGCGCAGG 60
 Db 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTGCTCGGGCGCTGTGGCGCAGG 60
 QY 61 ACAAGAGACACCCCGGATTCCTGATCCCAAGACTTTGCAACAGTTTTACCATCTGGGT 120
 Db 61 ACAAGAGACACCCCGGATTCCTGATCCCAAGACTTTGCAACAGTTTTACCATCTGGGT 120
 QY 121 GGGTCACTGGCTCGAGGGGGAACAGCTTCAAGCATGGCAATGAATCTACATGCTC 180
 Db 121 GGGTCACTGGCTCGAGGGGGAACAGCTTCAAGCATGGCAATGAATCTACATGCTC 180
 QY 181 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
 Db 181 CCTCAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240
 QY 241 AGCAGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCACCTCCTC 300
 Db 241 AGCAGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCACCTCCTC 300
 QY 301 TTTTCATGAATGCTTATACCATGAGGAGCTGGCGCAGTGATTCATACCCACTCTTAAAG 360
 Db 301 TTTTCATGAATGCTTATACCATGAGGAGCTGGCGCAGTGATTCATACCCACTCTTAAAG 360
 QY 361 CTGCTGTGATGGCTACCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
 Db 361 CTGCTGTGATGGCTACCTCTGTTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
 QY 421 TGATCAAGGAATTAAGGAATGATCTCAGGAGGCTTATACAGATACGATGATGTAG 480
 Db 421 TGATCAAGGAATTAAGGAATGATCTCAGGAGGCTTATACAGATACGATGATGTAG 480
 QY 481 TGGTACCTATTATTGAGAACACTCCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
 Db 481 TGGTACCTATTATTGAGAACACTCCTGAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
 QY 541 CCATGAATGAGTACCCAGACTCTGTCGGTCTTCTGCGGCTCATGGGCTGACGTGT 600
 Db 541 CCATGAATGAGTACCCAGACTCTGTCGGTCTTCTGCGGCTCATGGGCTGACGTGT 600
 QY 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
 Db 601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
 QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720
 Db 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAATG 720
 QY 721 GAATTTGTAAAGCAAGTGTGCTTAAAGCAATCCAAACATCAAACTCAATATG 780
 Db 721 GAATTTGTAAAGCAAGTGTGCTTAAAGCAATCCAAACATCAAACTCAATATG 780

XX DE DNA encoding antigen recognised by Ab capable of inducing G-CSF activity.
XX KW Antigenic protein; antibody; granulocyte colony stimulating factor;
XX KW G-CSF; cancer therapy; bone marrow suppression; human; ds.
XX OS Homo sapiens.
XX PN WO200060075-A1.
XX PD 12-OCT-2000.
XX PF 31-MAR-2000; 2000WO-JP002080.
XX PR 01-APR-1999; 99JP-00095092.
XX PA (NIBS) JAPAN TOBACCO INC.
XX PI Sha S, Aoki Y, Nishi Y;
XX DR WPI; 2001-024452/03.
XX DR P-PSDB; AAY85636.
XX PT Gene encoding an antigen recognizing an antibody which induces
PT granulocyte colony stimulating factor (G-CSF) expression for gene therapy
PT and treatment of G-CSF associated disorders e.g. the side effects of
PT cancer therapy.
XX Claim 4; Page 50-52; 58pp; Japanese.
XX The present invention relates to a gene encoding an antigenic protein
CC recognised by an antibody or its fragments which can induce the
CC production of granulocyte colony stimulating factor (G-CSF). Also
CC included in the invention are partial sequences of the gene, antibodies
CC recognising all or part of the antigenic protein, expression vectors
CC containing the gene and host cells transformed by the vector. The gene is
CC used for gene therapy, and compounds identified by screening using the
CC gene sequence are used for treatment and prevention of disorders
CC associated with G-CSF expression such as the side effects of cancer
CC therapy (including bone marrow suppression). The present sequence
CC represents the human gene of the invention
XX
SQ Sequence 924 BP; 291 A; 166 C; 225 G; 242 T; 0 U; 0 Other;
Query Match 70.9%; Score 595.4; DB 5; Length 924;
Best Local Similarity 84.7%; Pred. No. 3.5e-127;
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;
4 CCATGCTGCTGCTCAAGCT---CAAGGAGACTGTTGCTCGCGCGCGTGTGGCGCGCAGG 60
44 CCATGCTGCTGCTGATGCTTGGGAGGAGACTGTTGTTCCCGGAGATGCGGCGCGCAGG 103
61 ACAAGAGACACCCCGATTCTGATCCCAAGACTTTCGCAACAGTTTACCATCTGGGCT 120
104 ACAAGGAGCATCCAGATACCTGATCCAGAACTTTGCAAAACAGTTTACCATTTAGGCT 163
121 GCGTCACTGCACTGGAGGGGATGCTTGAAGCATGCAATGAATCTACATTGCTC 180
164 GGGTCACTGGGACTGGAGGAGGAATAGCTTGAAGCATGGGATGAATCTACATTGCTC 223
181 CCTCAGCGGTGCAAAAGGAGCGCATTCAGCGAGAGACATGTTTGTGTGACATTATG 240
224 CTTTCAGGAGTGCAAAAGGAGCAATTCAGCTGAGACATGTTTGTGATGATATAATG 283
241 AGCAGGACATAGCGGCGCTCCAGCATCTTAAGAGCTGAAAAAGGCGAGTGCCTCCTC 300
284 AAAAGGACATAAGTGGACCTTCGCCATCTGAGAGAGCTTAAAAAAGGCGAGTGTACTCCTC 343
301 TTTTTCATGATGCTTATACATGAGAGAGCTGGCGAGTGATTCATACCCACTCTAAAG 360
344 TTTTTCATGATGCTTATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
361 CTGCTGTGATGGCTACCCCTTCTGTTTCCAGGACAGGAGTTTAAAAATTACATCAAGAGA 420

404 CTGCTGTGATGCCACCTTCTCTTCCAGGACGGAGTTTAAAAATTACATCAAGAGA 463
421 TCATCAAAAGGAATAAGAAATGTACTCTCAGGAGGCTATTACAGATACGATGATGTTAG 480
464 TGATTAAGGAATAAGAAATGTACTCTCAGGAGGCTATTATAGATATGATGTTAG 523
481 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGATCTCAAAGAAAGATGGCTCATG 540
524 TGGTACCTATTATTGAGAACACTCTCTGAAGAGAGATCTCAAAGAAAGATGGCTCATG 583
541 CCATGAATGAGTACCCAGACTCTCTGTCGGTCTTGTTCGGGCTCATGCGGTGCTGCT 600
584 CAATGAATGATACCCAGACTCTCTGTCAGTACTGCTCAGACGTCATGAGTATATGTT 643
601 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACTCTGTTTGACA 660
644 GGGGGGAAACATGGGAGAAAGCCAAACCAATGTGTGAGTGTATGACTATTATTGATA 703
661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAAAATG 720
704 TTGCGGTATCAATGAAGAAAGTAGGACTTGTATCTTCCAGCTCCAGTTGGAGAAAAATG 763
721 GAATGTGTGACCAAGTGTGATGCTTACGATCTCCAAACAATAAAC-AACTCAATTAT 779
764 GAATGTGTGACCAAGTGTGATGCTTACGATCTCCAAACAATAAAC-AACTCAATTAT 823
780 GCCTTAATAATAAATCAAGCTGCTTGTAAAAAATAAATAA 816
824 TATTTAATGAAGCTATTATTTTAAATGATGATA 860
RESULT 6
ABK47967
ID ABK47967 standard; cDNA; 1136 BP.
XX AC ABK47967;
XX DT 02-JUL-2002 (first entry)
XX DE Human cDNA encoding G-CSF-inducible antibody binding protein, MMR19.
XX KW Human; granulocyte-colony stimulating factor; G-CSF; ss; gene; MMR19;
XX KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 46..774
XX FT /*tag= a
XX FT /product= "G-CSF-inducible antibody binding protein"
XX PN WO200226978-A1.
XX PD 04-APR-2002.
XX PF 27-SEP-2001; 2001WO-JP008446.
XX PR 27-SEP-2000; 2000JP-00294191.
XX PA (NIBS) JAPAN TOBACCO INC.
XX PI Sha S, Mukai H, Aoki Y, Nishi Y;
XX DR WPI; 2002-340016/37.
XX DR P-PSDB; AAU77178.
XX PT Gene encoding protein binding to antibody having granulocyte-colony
XX PT stimulating factor (G-CSF) inducing activity, useful for screening
XX PT potential drugs treating G-CSF associated diseases.
XX PS Claim 4; Page 94-96; 103pp; Japanese.

XX The invention relates to a mouse or human gene (MMR19) encoding a protein
CC which binds to antibodies or their fragments which induce granulocyte-
CC colony stimulating factor (G-CSF) secretion. The genes and proteins of
CC the invention are used in diagnosis, treatment and prevention of diseases
CC associated with G-CSF, including infections and neutrophil deficiency
CC disease. This sequence represents a cDNA encoding a human G-CSF-inducible
CC antibody binding protein, MMR19
XX
XX Sequence 1136 BP; 371 A; 188 G; 248 G; 329 T; 0 U; 0 Other;

Query Match 70.9%; Score 595.4; DB 6; Length 1136;
Best Local Similarity 84.7%; Pred. No. 3.7e-127;
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGTCAAGCT---CAAGGAGACTGTGTCGCGCGCTGCGCGGCGAGG 60
Db |||||
QY 44 CCATGCTGGCTGTGATGCTTGGGAGGAGACTGTGTTCCCGGAGATCGGCGCGCAGG 103
Db |||||
QY 61 ACAAGGAGCACCCTGATTCCTGATCCCAAGACTTTGCAACAGTTTACCATCTGGCT 120
Db |||||
QY 104 ACAAGGAGATCCAGATACCTGATCCCAAGACTTTGCAACAGTTTACCATTTAGCT 163
Db |||||
QY 121 GGGTCACTGGCACTGAGGGGAAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180
Db |||||
QY 164 GGGTCACTGGCACTGAGGGGAAATAGCTTGAAGCATGGCAATGAAATCTACATGCTC 223
Db |||||
QY 181 CCTCAGGCTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATG 240
Db |||||
QY 224 CTTTCAAGGAGTGAAGGAGGAAATTTTTCAGCCGAGACATGTTGTTTATGATATAATG 283
Db |||||
QY 241 AGCAGGACATAAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCATCTCCTC 300
Db |||||
QY 284 AAAAGGACATAGTGACCTTCGCCATCGAGAGCTTAAAAAGCCAGTGTACTCCTC 343
Db |||||
QY 301 TTTTCAATGATGTTATACATGAGAGAGCTGGCGAGTGAATTCATCCACTCTAAAG 360
Db |||||
QY 344 TTTTCAATGATGTTATACATGAGAGAGCTGGCGAGTGAATTCATCCACTCTAAAG 403
Db |||||
QY 361 CTGCTGTGATGCTACCTCTGTTTCCAGGACAGCTTTAAATTTACACATCAAGAGA 420
Db |||||
QY 404 CTGCTGTGATGCTACCTCTGTTTCCAGGACAGCTTTAAATTTACACATCAAGAGA 463
Db |||||
QY 421 TGATCAAGGAAATGAAGAAATGTACCTCAGGAGGCTTATACAGATACGATGATGTTAG 480
Db |||||
QY 464 TGNATAAGGAAATGAAGAAATGTACTTCCGAGGGTATTATAGATGATGATGTTAG 523
Db |||||
QY 481 TGGTACTTATTATGAGAACTCTGAGAGAGAGATCTCAAAGAAAGATGGCTCATG 540
Db |||||
QY 524 TGGTACTTATTATGAGAACTCTGAGAGAGAGATCTCAAAGAAAGATGGCTCATG 583
Db |||||
QY 541 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db |||||
QY 584 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 643
Db |||||
QY 601 GGGGAGAAACATGGGAGAAACAAACCATGCTGATGATGATGATGATGATGATGATGATG 660
Db |||||
QY 644 GGGGAGAAACATGGGAGAAACAAACCATGCTGATGATGATGATGATGATGATGATGATG 703
Db |||||
QY 661 TTGCTGCTTCCATGAGAGAGATGGGACTCGATCCAAACAGCTCCCGATGGGAGAAATG 720
Db |||||
QY 704 TTGCTGCTTCCATGAGAGAGATGGGACTCGATCCAAACAGCTCCCGATGGGAGAAATG 763
Db |||||
QY 721 GAATTTGTTAAGCAAGTGAATGCTTGAAGCATCTCCAAACATTAAC-AAACTCAATTAT 779
Db |||||
QY 764 GAATTTGTTAAGCAAGTGAATGCTTGAAGCATCTCCAAACATTAAC-AAACTCAATTAT 823
Db |||||
QY 780 GCCTTAATAAATCAAGCTGCTTTTAAAAAATAA 816
Db |||||
QY 824 TATTTAATGAAGCTATTTTAAATGAATTCAAA 860
Db |||||

AAAX05748/c
ID AAAX05748 standard; cDNA; 1190 BP.
XX
AC AAAX05748;
XX
DI 28-APR-1999 (first entry)
XX
DE Nucleotide sequence of human HFI2G53.
XX
KW HFI2G53; human; inflammatory disease; infection; HIV-1; HIV-2; cancer;
KW HIV-associated cachexia; immunodeficiency disorder; septic shock; pain;
KW Parkinson's disease; cardiovascular disease; psychotic; neurological;
KW Huntington's disease; Gilles de la Tourette's syndrome; gene mapping;
KW mental retardation; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 154..879
FT /*tag= a
FT /note= "the specification indicates that this region
FT encodes the human HFI2G53 protein (AAW94762); the
FT translated product of this sequence does not match the
FT protein sequence AAW94762"
XX
PN EP892050-A2.
XX
PD 20-JAN-1999.
XX
PF 17-FEB-1999; 98EP-00301168.
XX
PR 08-JUL-1997; 97US-0051937P.
PR 17-OCT-1997; 97US-00953494.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Demarini D;
XX
DR WPI; 1999-083567/08.
DR P-PSDB; AAW94762.
XX
PT New HFI2G53 polypeptide and polynucleotide - useful as diagnostic
PT reagents and for prevention and treatment of inflammatory diseases,
PT cancer and Parkinson's disease.
XX
PS Claim 2; Page 6-7; 22pp; English.
XX
CC This represents the nucleotide sequence of human HFI2G53. Host cells
CC containing an expression system comprising the HFI2G53 nucleic acid are
CC used for the recombinant production of the protein. HFI2G53 polypeptides
CC and polynucleotides are useful for diagnosing diseases related to over or
CC underexpression of HFI2G53 protein. The HFI2G53 polypeptides can be used
CC to screen for agonists and antagonists which can be used in treatment to
CC activate or inhibit HFI2G53 activity. Gene therapy may also be used to
CC affect endogenous polypeptide production, using HFI2G53 polynucleotides
CC and retroviral vectors. HFI2G53 antibodies are useful for inducing an
CC immune response to immunise and prevent diseases, and for isolating
CC HFI2G53 clones or purifying the polypeptide by affinity chromatography.
CC HFI2G53 polypeptides can be administered directly or as a vaccine to
CC inoculate against disease. Diseases prevented, diagnosed or treated
CC include inflammatory diseases such as Adult Respiratory Disease Syndrome,
CC rheumatoid arthritis, osteoarthritis, inflammatory bowel disease, asthma,
CC psoriasis, dermatitis, allergies; infections including bacterial, fungal,
CC protozoan and viral, particularly HIV-1 and -2; HIV-associated cachexia
CC and other immunodeficiency disorders; septic shock; injury; pain; cancers
CC including testicular cancer; anorexia; bulimia; Parkinson's disease;
CC cardiovascular disease including restenosis, atherosclerosis, acute heart
CC failure, myocardial infarction, hypertension, hypertension; urinary
CC retention; angina pectoris; ulcers; benign prostatic hypertrophy; and
CC psychotic and neurological disorders (anxiety, schizophrenia, delirium,
CC manic depression, dementia, severe mental retardation) and dyskinesias,
CC such as Huntington's disease or Gilles de la Tourette's syndrome. The
CC HFI2G53 polypeptide is also useful for mapping the gene to a chromosome,
CC

CC allowing gene inheritance to be studied through linkage analysis

XX Sequence 1190 BP; 317 A; 287 C; 240 G; 346 T; 0 U; 0 Other;

Query Match 70.9%; Score 595.4; DB 2; Length 1190;

Best Local Similarity 84.7%; Pred. No. 3.8e-127;

Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;

QY 4 CCATGCTGCTGCTCAAGCT---CAAGGAGACTGTTCTCGCGCGCTGTGGCGCGCAGG 60
DB 995 CCATGCTGCTGCTGCTGCTGGAGGAGACTGTTGTTCCGGAGATGCGGCGCGCAGG 936
QY 61 ACAAGAGACCCCGGATCTCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 120
DB 935 ACAAGAGACCTCAAGATACCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 876
QY 121 GGCTCACTGCACTGGAGGGGATCAGCTTGAGCATGCAATGAATCTACATGCTC 180
DB 875 GGCTCACTGCACTGGAGGGGATCAGCTTGAGCATGCAATGAATCTACATGCTC 816
QY 181 CCTCAGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATG 240
DB 815 CTTGAGAGTGCAAAAGGAGCGCATTCAGCTGAGCATGTTGTTTATGATATAATG 756
QY 241 AGCAGGACATAAGCGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300
DB 755 AAAAGGACATAAGTGGACCTTGGCCATCGAAGAGCTTAAAAAGGCGCACTGTA 696
QY 301 TTTTCATGATGCTTATACATGAGAGAGTGGCGAGTGATTCATACCCACTCTAAG 360
DB 695 TTTTCATGATGCTTATACATGAGAGAGTGGCGAGTGATTCATACCCACTCTAAG 636
QY 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTTACACATCAAG 420
DB 635 CTGCTGTGATGGCGACCCCTTCTTCCAGACGGAGTTTAAATTTACACATCAAG 576
QY 421 TGATCAAGGAATTAAGAAATGTACCTCAGAGGCTATTACAGATAGATGATGTTAG 480
DB 575 TGATCAAGGAATTAAGAAATGTACTTCCGAGGCTATTATAGATATGATGATGTTAG 516
QY 481 TGGTACCTATTATGAGAACTCTCTGAGAGAGGATCTCAAGAGAGGATGGCTCATG 540
DB 515 TGGTACCTATTATGAGAACTCTCTGAGAGAGGATCTCAAGAGAGGATGGCTCATG 456
QY 541 CCATGATGATGATCCAGACTCTGTGCGGTTCTGTCCGCGCTCATGGGCTGTACGTG 600
DB 455 CAATGATGATGATCCAGACTCTGTGCGGTTCTGTCCGCGCTCATGGGCTGTACGTG 396
QY 601 GGGAGAGAAATGGGAGAGAAAGCAAAACCATGTGTGATGTTATGACTCTGTTTGACA 660
DB 395 GGGGAGAAATGGGAGAGAAAGCAAAACCATGTGTGATGTTATGACTCTGTTTGATA 336
QY 661 TTGCTGTCTCCATGAGAGATGGGATCGATCCATCAACAGCTCCAGTGGGAGAAATG 720
DB 335 TTGCTGTCTCCATGAGAGATGGGATCGATCCATCAACAGCTCCAGTGGGAGAAATG 276
QY 721 GAATGTGTAGCAAGTGGATGCTTGAAGTCTTCCAACTAAATAAAC-AAACTCAATTAT 779
DB 275 GAATGTGTAGCAAGTGGATGCTTGAAGTCTTCCAACTAAATAAAC-AAACTCAATTAT 716
QY 780 GCCTTAATAAATCTCAGCTCTTTTAAAAAATAAATAA 816
DB 215 TATTAAATGAAGCTATTTTAAAAATGAATTGAAA 179

RESULT 8

ID ABK12566

ABK12566 standard; cDNA; 1226 BP.

XX AC ABK12566;

XX DT 18-JUN-2002 (first entry)

XX XX

DNA encoding cell differentiation stimulator associated protein #1.

Cartilage cell differentiation stimulator; osteopathic;

Membrane-bound transferrin-like protein; Mtf-BP; concanavalin A; ConA;

Membrane bound type transferrin-like protein; Mtf; cartilage disorder;

bone metabolism disease; cell differentiation; cell growth;

extracellular matrix related disease; gene; ss; human.

Homo sapiens.

Location/Qualifiers

Key 78, 806

CDS /tag= a

/product= "Cell differentiation stimulator associated

protein #1"

JP2002C20311-A.

23-JAN-2002.

07-JUL-2000; 2000JP-00206566.

07-JUL-2000; 2000JP-00206566.

(KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.

WPI; 2002-287405/33.

P-PSDB; AAU78360.

A cartilage cell differentiation stimulator useful in the diagnosis of

biophylaxis, cell differentiation, cell growth and construction of

extracellular matrix related diseases.

Claim 5; Page 8; 17pp; Japanese.

The invention describes a cartilage cell differentiation stimulator

(containing a membrane-bound transferrin-like protein (Mtf-BP) and a

membrane bound type transferrin-like protein (Mtf)) and an animal-derived

concanavalin-like drug. The cartilage differentiation stimulator can be

used in diagnosis, prevention and treatment of cartilage and bone

metabolism diseases. They can also be used for diagnosing biophylaxis,

cell differentiation, cell growth and construction of extracellular

matrix related diseases. Mtf-BP strongly stimulates differentiation of

cartilage cells and exhibits similar action mechanism with that of plant

derived ConA. This sequence represents a cartilage cell differentiation

stimulator associated polypeptide described in the invention

Sequence 1226 BP; 382 A; 212 C; 274 G; 358 T; 0 U; 0 Other;

Query Match 70.7%; Score 593.8; DB 6; Length 1226;

Best Local Similarity 84.6%; Pred. No. 8.9e-127;

Matches 691; Conservative 0; Mismatches 122; Indels 4; Gaps 2;

QY 4 CCATGCTGCTGCTCAAGCT---CAAGGAGACTGTTCTCGCGCGCTGTGGCGCGCAGG 60

DB 76 CCATGCTGCTGCTGCTGCTGGAGGAGACTGTTGTTCCCGAGATGGCGCGCAGG 135

QY 61 ACAAGAGACCCCGGATCTCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 120

DB 136 ACAAGAGACTCCAGATACCTGATCCAGAACTTTGCAAAAGTTTACATCTGGGCT 195

QY 121 GGCTCACTGCACTGGAGGGGATCAGCTTGAGCATGCGCAATGAATCTACATGCTC 180

DB 196 GGCTCACTGCACTGGAGGGGATCAGCTTGAGCATGCGCAATGAATCTACATGCTC 255

QY 181 CCTCAGCGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGACATTAATG 240

DB 256 CTTGAGAGTGCAAAAGGAGCGAATTCAGCTGAGACATGTTGTTGTGATATAATG 315

QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTAAGAGCTGAAAAAGGCGAGTGCCTCCTC 300

DB 316 AAAAGGACATAAGTGGAGCTTCCCATCGAAGAGCTTAAAAAGGCGAGTGTACTCCTC 375

Db 122 CCATGAAGAGATGGGCTGGATCCCAACGCGAGTTCCTCCAGTTGGAGAACATGGATCGTAT 63
QY 730 AAGCCAGTGGATGCTTACGATCTCCACAAATAAACAACACTCAATTATGCTTAAATA 789
Db 62 AAGCCAGTGGAGCGCTTAAAGATCTCCAAAATAAACAACACTCAATTATGCTTAAATA 3
QY 790 AA 791
Db 2 AA 1
RESULT 10
ID ABX63252/c
AC ABX63252;
XX 25-FEB-2003 (first entry)
DT Human cDNA #252 differentially expressed in activated vascular tissue.
DE Human; gene; sr; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW gene therapy; vascular disease; cancer; coronary; artery disease;
KW hypertension; diabetes; pre-eclampsia; restenosis;
KW ischaemia-reperfusion injury; stroke.
XX Homo sapiens.
OS US2002137081-A1.
PN 26-SEP-2002.
XX 08-JAN-2002; 2002US-00044090.
XX 28-JUL-2000; 2000US-0222469P.
PR 08-JAN-2001; 2001US-0260483P.
XX (BAND/) BANDMAN O.
PA Bandman O;
PI WPI; 2003-110597/10.
DR Combination for diagnosing, staging, treating, or monitoring the
PT progression of treatment of a vascular disease, e.g. atherosclerosis,
PT comprises several cDNAs that are differentially expressed in activated
PT vascular tissue.
XX Claim 1; Page; 18pp; English.
XX This invention relates to a combination comprising several cDNAs that are
XX differentially expressed in activated vascular tissue. The invention also
XX discloses a high throughput method for detecting differentially expressed
XX cDNAs in a sample. The cDNAs of the invention may have
XX antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
XX gynaecological; vasotropic and cerebroprotective activities and may be
XX used in gene therapy. The cDNAs of the invention may be used in a high-
XX throughput methods for detecting differential expression of one or more
XX cDNAs in a sample, or screening several molecules or compounds to
XX identify a molecule or compound that specifically binds a cDNA of the
XX invention. A protein encoded by the cDNA may be used to screen several
XX molecules or compounds to identify a ligand that specifically binds to
XX the protein, or to produce or purify an antibody to the protein that can
XX be used to detect a protein in a sample or purify a natural or
XX recombinant protein from a sample. The nucleotides may be useful for
XX diagnosing, staging, treating, or monitoring the progression of treatment
XX of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
XX disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion
XX injury, restenosis, or stroke. The cDNAs can also be used for large-scale
XX genetic or gene expression analysis of several new nucleic acid
XX molecules. Antibodies to the proteins encoded by the cDNAs are useful for

CC diagnosing pre-pathologic disorders, and chronic or acute diseases
CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC http://seqdata.uspto.gov/sequence.html?docID=20020137081
XX
SQ Sequence 3764 BP; 1175 A; 758 C; 777 G; 1053 T; 0 U; 1 Other;
Query Match 47.6%; Score 399.8; DB 7; Length 3764;
Best Local Similarity 79.9%; Pred. No. 4.6e-82;
Matches 671; Conservative 0; Mismatches 143; Indels 26; Gaps 16;
QY 4 CCATGTCTGGCTGTCAAGCTCAAGGAG-----ACTGTGTCTCGCGCGCTGTGCGGCGAG 59
Db 1182 CCATGTCTGGCTGTGATGCTCGGGAGGAGACTGGTGTTCCTCCGAGATGCGCGCGAG 1123
QY 60 GACAAGGAGACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGC 119
Db 1122 GACAAGGAGATCCAGATACCTGATCCAGAACTTTGCAACAGTTTACCATTTAGGC 1063
QY 120 TGGGTCTAC--TGSCACTGGAGGGGGAATCAGCTTTGAAGCATGCG--AATGAAATCTCAT 176
Db 1062 TGTGTCTACCTGTGCTGCTGAGGAGGAACTTAGCTTTGAAGCATGCGGATTTAAATCTCAT 1003
QY 177 GCTCCCTCAGGCGTGCMAAAGGAGCGCATTCAGCCAGAAAGACATGTTTGTGT--GTGACAT 235
Db 1002 GCTCCTTCAGGAGTGCMAAAGGAGCGAATTCAGCTTGAAGCATGTTTGTGTATGATAT 943
QY 236 TAAATGAGGAGCATAAAGCGGG--CTTCAGCATCTTAAGAGCT--GAAATAAGCGAGTGC 293
Db 942 AAATGAAAGGACATAAGTGGGACCTTCGCGATCGAAGAGCTTAAATAAGCGAGTGT 883
QY 294 ACTCTCT--TTTTCAATCAATGCTTATACCATGAGAGGCTGGCGGAGTGTATCATACCCA 352
Db 882 ACTCTCTTTTTTCATGGATGCTTACCAATGAGAGGAGGAGTGTGAGTGTATCATACCCA 823
QY 353 CTCTAAAGCTGTGTGATGGTACCCTTCT--GTTCCAGGACAGGAGTTTAAATAATACAC 411
Db 822 CTCTAAAGCTGTGTGATGGGACCTTCTCTTTCCAGGACGGAGTTTAAATAATACAC 763
QY 412 ATCAAGAGATGATCAAAAGGAATAAGGAATGTACTCTCAGAGGCTATTTACAGATACATG 471
Db 762 ATCAAGAGATGATAAAAGGAATAAGGAATGTACTCTCCGAGGGGTATTTAGATATGATG 703
QY 472 ATATGTAGTGGT--ACTTATTTAGTACACTCTCTGA--AGAGAAGATCTTCAAGAAAG 529
Db 702 ATATGTAGTGGTGGGCTCCCTATTTAGATTTCCACCTGAGGAGGAGAAAGACCTCAAGATTG 643
QY 530 GA--TGGCTCATGCCATGAATG--AGTACCCAGACTCC---TGTGGGGTTCCTTGTCCGGC 582
Db 642 GAATGGCTCATGCAATGAATGAATGAATACCCAGACTCCCTGTNCCAGTACTGTGTAGAC 583
QY 583 GT--CATGGGTGTACGTGTGGGAGAAACATGGGAGAAAG--CAATAACCATGTGTGAGT 639
Db 582 GTCCATGAGTATATGTGTGGGGGAAACCATGGGAGAAAGGCCCAAAACCATGTGTGAGT 523
QY 640 GTTATGACTACTGTGTGTGACATTTGCTCTCCATGAAGAGATGGGACTCGATCCCAAC 699
Db 522 GTTATGACTATTTTATTTGATTTATTCGCGTATCAATGAAGAAAGTAGGACTTGTATCTTTCAC 463
QY 700 AGCTCCAGTTGGAGAAATGGAATTTGTGTGAAGCAAGTGTGATGCGCTCAAGCATCTCCAAC 759
Db 462 AGCTCCAGTTGGAGAAATGGAATTTGTCTAGCCAAAGAAAGTCTAATTTATATACAGA 403
QY 760 AATFAAAC--AAACTCAATATGCTTTAATAAATACTAGCTGCTTTTAAAAAATAAAAA 817
Db 402 GATAAGCTAAACGTTTAAATTTATTTTAAATGAAGCTATTTTAAAAATGAATTTGAAA 343
RESULT 11
AAZ16856

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randolph Z, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;

DR WPI; 1999-494092/41.
XX Novel human genes and their expression products which are differentially
PT expressed in different cell types.
PS Claim 1; Page 2200; 2479pp; English.
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA212532 to AA217779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one of
CC the 5248 polynucleotide sequences given in AA212532 to AA217779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists
XX Sequence 778 BP; 244 A; 129 C; 150 G; 241 T; 0 U; 14 Other;
SQ Query Match 42.1%; Score 354; DB 2; Length 778;
Best Local Similarity 81.1%; Pred. No. 1.1e-71;
Matches 423; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
Qy 298 CTCTTTTCATGATGTTTATACCATGAGAGAGAGTGGCGGAGTTCATACACCTCTA 357
Db 46 CTTTNTGAGGATCCCATCGATTCGAGAGAGAGAGTGGCGGAGTTCATACACCTCTA 105
Qy 358 AAGCTGCTGTGAGGCTACCTCTGTTTCCAGACAGGAGTTTAAATTCACATCAAG 417
Db 106 AAGCTGCTGTGAGGCTACCTCTGTTTCCAGACAGGAGTTTAAATTCACATCAAG 165
Qy 418 AGATGATCAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATACGATGATGT 477
Db 166 AGATGATCAAGGAATAGGAATGTACCTCAGGAGGCTATTACAGATGATGATGT 225
Qy 478 TAGTGGTACCTATTATTGAGACACTCCTGAGAGAGAGATCTCAAGAGAGGATGGCTC 537
Db 226 TAGTGGTACCTATTATTGAGACACTCCTGAGAGAGAGATCTCAAGAGAGGATGGCTC 285
Qy 538 ATGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 597
Db 286 ATGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 345
Qy 598 TGTGGGAGAGAAATGGAG 657
Db 346 TGTGGGAGAGAAATGGAG 405
Qy 658 ACATTGCTGTCTCCATGAAGAAGATGGAGCTCGATCCAAACACAGCTCCCGATTGGAGAA 717
Db 406 ATATTGCGGTATCAATGAAGAAGATGGAGCTCGATCCAAACACAGCTCCCGATTGGAGAA 465
Qy 718 ATGGAATGTGTAGCCAGAGTGGATGCTTAGCATCTCC--AACAAATAAACAACATCAA 775
Db 466 ATGGAATGTGTAGCCAG 525
Qy 776 TTATGCTTAAATAAATCAAGCTGCTTTTAAAAA 816
Db 526 TTATTAATTAATGAAGCTATTTTAAATGAATTGAAA 566

RESULT 13

AA215675
ID AA215675 standard; cDNA; 851 BP.
XX AA215675;
AC AA215675;
XX 12-OCT-1999 (first entry)
DT Human gene expression product cDNA sequence SEQ ID NO:3144.
DE Human; gene; gene expression product; diagnosis; therapy; probe;
XX detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; es.
XX Homo sapiens.
OS WO9938972-A2.
PN 05-AUG-1999.
PD 28-JAN-1999; 99WO-US001619.
PF 28-JAN-1999; 98US-0072910P.
PR 24-FEB-1998; 98US-0075954P.
PR 31-MAR-1998; 98US-0080114P.
PR 03-APR-1998; 98US-0080515P.
PR 03-APR-1998; 98US-0080666P.
PR 21-OCT-1998; 98US-0105234P.
PR 28-OCT-1998; 98US-0105877P.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Ckvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX WPI; 1999-494092/41.
XX Novel human genes and their expression products which are differentially
PT expressed in different cell types.
PS Claim 1; Page 1511; 2479pp; English.
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA212532 to AA217779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one of
CC the 5248 polynucleotide sequences given in AA212532 to AA217779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists
XX Sequence 851 BP; 266 A; 143 C; 162 G; 243 T; 0 U; 37 Other;
SQ Query Match 41.5%; Score 348.2; DB 2; Length 851;
Best Local Similarity 81.1%; Pred. No. 2.4e-70;
Matches 415; Conservative 0; Mismatches 96; Indels 1; Gaps 1;
Qy 308 GATGCTTATACCATGAGAGAGAGTGGCGGAGTTCATACACCTCAAGCTGTGT 367

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

WPI: 2001-656860/75.

AA New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell
PT interactions.

PS Claim 1: SEO ID NO 3484: 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from prosochilla. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU161176-ABU130511), expressed DNA sequences (ABU101840-ABU161175) and the encoded proteins (ABR57737-ABR72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published pct sequences

Sequence 796 BP: 194 A: 214 C: 220 G: 168 T: 0 U: 0 Other: 0

Query Match 30.5%; Score 256; DB 4; Length 796;

Query Match 50.5%; Score 250; Length 150
Best Local Similarity 62.5%; Pred. No. 3.7e-49;

```

Matches 400; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

```

QY 65 GGAGCACCCCGATTCTGTATCCAGAACTTTGCAAACAGTTTACCACATCTGGGCTGGGT 124

Db
140 GGAGCATCCTCGCCACTTGATTCCCTCGCTATGCAGGCAATTCTATCATTTGGGATGGGT 199

125 CACTGGCACTGGAGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATTGCTCCCTC 184

200 GACCGGCACAGGAGGTGGCATTAAGTACAACGATGAGATCTACATAGCACCGTC 259

185 AGCGTGCGGAAAGGAGCGCATTCAGCCAGAAAGACATGTTTGTGTGTGACATTAATGAGCA 244

260 CCGCCCTCCGACCGACCGGAAATGCGAGCGGAGGATCTCTTCGTGCAGGATATAACGGCAA 319

304

379

[illegible]

7

2000

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

DD CAHGGGCGTCTACGAAGAGCCGACCAAGCAGATTAATTGCCTAAGCACAAACCATACTCTTT
300

QY 485 ACC TAT TAT T GAG A CAC TCC TGA AGA GAG A GGC ATC CAA AGA GAG A GGC ATC CCA TCC

DB 560 ACCGATCATCGAGAACACACCTTTTGAAAGCGGACCTGGCCGACAGTAGTAGCGCCGCGCAI 51

QY
545 GAATGAGTACCCAGACTCCCTGTGCGGTCTCTGTCGGCGTCAAGGGGTGACGIGGGG 60

Db 620 GATGGAGTATCCGGGCTGCAGTGGCATCCCTGGTTCGACGACACGGCGTCTACGTTTGGGG 673

QY 605 AGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGC 664

Db 680 ACAGAACTGGAGAAGGCCAAACCATGTCGGAATGCTATGACTATCTCTTCTCCATTGC 739

QY 665 TGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTC 704

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 22:00:56 ; Search time 96 Seconds
(without alignments)
4855.820 Million cell updates/sec

Title: US-09-937-905-1
Perfect score: 840
Sequence: 1 gaaccatgtctggtctca.....aaaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	323	38.5	409	4	US-09-621-976-9280
2	143	17.0	229	1	US-08-840-683-4
3	143	17.0	229	2	US-08-555-722-4
4	143	17.0	229	3	US-09-384-301-4
5	53.6	6.4	1223	3	US-09-154-874-4
6	53.6	6.4	1223	4	US-08-931-668-4
7	53.6	6.4	1223	4	US-09-468-175-4
8	53	6.3	147	4	US-09-621-976-10254
9	53	6.3	7218	1	US-08-232-463-14
10	52.8	6.3	997	4	US-09-907-794A-376
11	52.8	6.3	997	4	US-09-905-125A-376
12	52.8	6.3	997	4	US-09-902-775A-376
13	52.2	6.2	147	4	US-09-621-976-10383
14	50.4	6.0	375	3	US-08-946-026-23
15	49.6	5.9	242	4	US-09-621-976-16320
16	48.8	5.8	1074	3	US-09-248-335-67
17	48.6	5.8	746	3	US-09-013-810-1
18	48.6	5.8	1558	1	US-08-455-550-7
19	48.2	5.7	350	1	US-08-171-385-14
20	48.2	5.7	350	3	US-08-361-441B-14
21	48.2	5.7	5555	1	US-08-484-438-3
22	47.8	5.7	1987	4	US-09-227-357-44
23	47.6	5.7	441	4	US-09-601-537-10
24	47.6	5.7	674	4	US-09-620-405B-465
25	47.6	5.7	674	4	US-09-433-826B-465
26	47.6	5.7	674	4	US-09-604-287A-465
27	47.6	5.7	674	4	US-09-834-759-465

28	47.6	5.7	1454	4	US-09-372-422A-19	Sequence 19, Appl
29	47.6	5.7	2186	4	US-09-360-545-66	Sequence 66, Appl
30	47.6	5.7	4121	4	US-09-601-537-9	Sequence 9, Appl
31	47.4	5.6	1474	3	US-08-821-994-64	Sequence 64, Appl
32	47.4	5.6	2422	4	US-09-369-247-53	Sequence 53, Appl
33	47.2	5.6	1323	3	US-09-413-452-3	Sequence 3, Appl
34	47.2	5.6	1323	3	US-09-413-068-3	Sequence 3, Appl
35	47.2	5.6	1323	4	US-09-403-345A-3	Sequence 4, Appl
36	47.2	5.6	1700	2	US-08-897-340-4	Sequence 4, Appl
37	47.2	5.6	1700	3	US-09-252-329-4	Sequence 4, Appl
38	47.2	5.6	2434	4	US-09-489-847-67	Sequence 67, Appl
39	47.2	5.6	3715	4	US-09-234-245-1	Sequence 1, Appl
40	47	5.6	1332	2	US-09-057-762-1	Sequence 1, Appl
41	47	5.6	1332	3	US-08-326-119A-1	Sequence 1, Appl
42	47	5.6	2628	1	US-08-143-219-1	Sequence 1, Appl
43	46.8	5.6	193	4	US-09-621-976-10543	Sequence 10543, A
44	46.8	5.6	1069	4	US-09-205-258-74	Sequence 74, Appl
45	46.8	5.6	3124	4	US-09-734-030-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-621-976-9280
; Sequence 9280, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 9280
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9280

Query Match	38.5%	Score 323;	DB 4;	Length 409;
Best Local Similarity	88.6%	Pred. No. 9.8e-84;		
Matches	350;	Conservative	0;	Mismatches 45;
			Indels	0;
			Gaps	0;
Qy	342	ATTGATACCCACTCTAAAGCTGCTGATGGCTACCTTCTGTTTCCAGGACAGGAGTTT	401	
Db	1	ATTGATACCCACTCTAAAGCTGCTGATGGCTACCTTCTGTTTCCAGGACAGGAGTTT	60	
Qy	402	AAATTTACATCAAGAGATGATCAAGGAATAGGAATGTACTCTCAGGAGGCTATTAC	461	
Db	61	AAATTTACATCAAGAGATGATCAAGGAATAGGAATGTACTCTCAGGAGGCTATTAT	120	
Qy	462	AGATACGATGATGTTAGTGTGCTACCTATTATTGAGAACACTCTCTGAGAGAGATCTC	521	
Db	121	AGATATGATGATGTTAGTGTGCTACCTATTATTGAGAACACTCTCTGAGAGAGATCTC	180	
Qy	522	AAAGAAAGGATGGCTCATGCCAATGAGTACCCAGACTCTCTGCGGTTCTTGTTCGG	581	
Db	181	AAAGATGAGATGGCTCATGCCAATGAGTACCCAGACTCTCTGCGGTTCTTGTTCGG	240	
Qy	582	CGTCATGGGTTGACGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT	641	
Db	241	CGTCATGGGTTGACGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT	300	
Qy	642	TATGACTACCTGTTTGCATGCTGTCTCCATGAAGAGATGGGACTCGATCCACACAG	701	
Db	301	TATGACTATTATTGATATTCGCTATCAATGAAGAAAGTAGGACTTGTATCTCTCACAG	360	
Qy	702	CTCCAGTGGAGAAATGGAATTGTCTAGCCAA	736	

Db 361 CTCCAGTTGGAGAAATGGAATTGTCTAAGCCAA 395

RESULT 2

US-08-840-683-4
; Sequence 4, Application US/08840683
; Patent No. 5821051
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; TELECOMMUNICATION INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-840-683-4

Query Match 17.0%; Score 143; DB 1; Length 229;
Best Local Similarity 85.2%; Pred. No. 7.3e-32;
Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108
Db 4 GCGGCCGACAGGACAGGACATCCAGTACTGATCCAGAAC-TTGCAACAGTTTT 62
QY 109 ACCATCTGGGCTGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168
Db 63 ACCATTTAGGCTGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120
QY 169 TCTACATTGCTCCCTCAGCGGTGCAAAAGGAGGCGCTCCAGCATCTAAGAAGCT 278
Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAGCAATTCAGCTTGAAGCATGTTTGT 179
QY 229 GTGACATTAAATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277
Db 180 GTGATATAAATGAAGAGGACATAAGTGGACCTTCGCCATCGAAGAGCT 228

RESULT 3

US-08-555-722-4
; Sequence 4, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-555-722-4

Query Match 17.0%; Score 143; DB 2; Length 229;
Best Local Similarity 85.2%; Pred. No. 7.3e-32;
Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 GTGGCGCGCAGGACAGGACCCCGATTCTGATCCAGAACTTTGCAAAACAGTTTT 108
Db 4 GCGGCCGACAGGACAGGACATCCAGTACTGATCCAGAAC-TTGCAACAGTTTT 62
QY 109 ACCATCTGGGCTGGTCACTGGCACTGGAGGGGAATCAGCTTGAAGCATGCAATGAAA 168
Db 63 ACCATTTAGGCTGGTCACTGGCACTGGAGGAGGAATTAGCTTGAAGCATGG--GTGAAA 120
QY 169 TCTACATTGCTCCCTCAGCGGTGCAAAAGGAGGCGCTCCAGCATCTAAGAAGCT 278
Db 121 TCTACATTGCTCCCTCAGG-ATGCAAAAGGAGCAATTCAGCTTGAAGCATGTTTGT 179
QY 229 GTGACATTAAATGACGAGGACATAAGCGGCGCTCCAGCATCTAAGAAGCT 277
Db 180 GTGATATAAATGAAGAGGACATAAGTGGACCTTCGCCATCGAAGAGCT 228

RESULT 4

US-09-384-301-4
; Sequence 4, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/384,301
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/555,722
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: NEP-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 229 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-384-301-4

Query Match 17.0%; Score 143; DB 3; Length 229;
Best Local Similarity 85.2%; Pred. No. 7,3e-32;
Matches 195; Conservative 0; Mismatches 30; Indels 4; Gaps 3;
QY 49 GTGGCGCAGACAGGAGGACCCCGCATTCCTGATCCAGAACTTTTGCAACACAGTTTT 108
DB 4 GGGCGCAGACAGGAGGACGATCCAGATCTGATCCAGAAC-TTGCACACAGTTTT 62
QY 109 ACCATCTGGCTGGGTCTACCTGGCACTGGAGGGGATCAGTTGAGCATGCGCAATGAAA 168
DB 63 ACCATTGAGCTGGGTCTACCTGGCACTGGAGGAGGATAGCTTGAAGCATGG-GTGA 120
QY 169 TCTACATTCCTCTCCTCAGCGGTGCAAAAGGAGGCACTTCAGCAGAGACATGTTGCT 228
DB 121 TCTACATTCCTCTCCTCAGG-ATGCAAAAGGAGCAATTCAGCCTGAAGACATGTTGTT 179
QY 229 GTGACATTAATGAGGAGGACATAAGCGGCTCCAGCATCTAAGAGCT 277
DB 180 GTGATATAATGAAGGACATAAGTGGACCTTGGCCATCGAAGAGCT 228

RESULT 5
US-09-154-874-4
Sequence 4, Application US/09154874
Patent No. 6054636
GENERAL INFORMATION:
APPLICANT: FADER, GARY MICHAEL
TITLE OF INVENTION: ISOFLAVONE BIOSYNTHETIC ENZYMES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,874
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/931,668
FILING DATE: SEPTEMBER 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1098-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 55..1223
FEATURE:
NAME/KEY: CDS
LOCATION: 55..990
US-09-154-874-4

Query Match 6.4%; Score 53.6; DB 3; Length 1223;
Best Local Similarity 73.9%; Pred. No. 1.2e-05;
Matches 68; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 749 GCATCTCCACAAATAAACAACTCAATATGCTTAATTAATAAATCAGCTCTTTTAA 808
DB 1113 GCATCTTGAATGTGAAGAGTTTAACCTTTATCATTAATTTAAGCTCTGTTTAA 1172
QY 809 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
DB 1173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1204

RESULT 6
US-08-931-668-4
Sequence 4, Application US/08931668
Patent No. 6521433
GENERAL INFORMATION:
APPLICANT: FADER, GARY M.
TITLE OF INVENTION: CDNA SEQUENCES FROM SOYBEAN THAT ENCODE
ACTIVITIES ASSOCIATED WITH ISOFLAVONE
BIOSYNTHESIS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,668
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1098
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 1223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 55..1223
FEATURE:
NAME/KEY: CDS
LOCATION: 55..990
US-08-931-668-4

Query Match 6.4%; Score 53.6; DB 4; Length 1223;
Best Local Similarity 73.9%; Pred. No. 1.2e-05;
Matches 68; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 749 GCATCTCAACAATAAACAACCAATTAATGCTTAATAAATAAATCAATGCTTTTAAA 808
DB 1113 GGATCTTGAATGTGAACGAGTTTAACTTTATCATTTAATTTAAGCTCTGTTTAAA 1172
QY 809 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
DB 1173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1204

RESULT 7
US-09-468-175-4
; Sequence 4, Application US/09468175
; Patent No. 6617493
; GENERAL INFORMATION:
; APPLICANT: FADER, GARY MICHAEL
; TITLE OF INVENTION: ISOFALVONE BIOSYNTHETIC ENZYMES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/468,175
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/931,668
; FILING DATE: SEPTEMBER 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARIAN, WILLIAM R.
; REGISTRATION NUMBER: 41,173
; REFERENCE/DOCKET NUMBER: BB-1098-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..1223
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..990

US-09-468-175-4

Query Match 6.4%; Score 53.6; DB 4; Length 1223;
Best Local Similarity 73.9%; Pred. No. 1.2e-05;
Matches 68; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 749 GCATCTCAACAATAAACAACCAATTAATGCTTAATAAATAAATCAATGCTTTTAAA 808
DB 1113 GGATCTTGAATGTGAACGAGTTTAACTTTATCATTTAATTTAAGCTCTGTTTAAA 1172
QY 809 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
DB 1173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1204

RESULT 8
US-09-621-976-10254
; Sequence 10254, Application US/09621976
; Patent No. 8639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10254
; LENGTH: 147
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10254

Query Match 6.3%; Score 53; DB 4; Length 147;
Best Local Similarity 61.0%; Pred. No. 5.8e-06;
Matches 86; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 730 AGCTCCCGAGTTGGAGAAAAATGGAATTTGTAAAGCAAGTGGATCGCTTAAGCATCTCCAAC 759
DB 1 AACTCTGTTTAGAGAGACATCTCAATGTTTAGTAAAGTACAAACACCATCAAAAAAAA 60
QY 760 AATAAACAACCACTCAATTAATGCTTTAAATAAATAAATCAATGCTTTTAAAAA 819
DB 61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120
QY 820 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
DB 121 AAAAAAAAAAAAAAAAAAAAA 141

RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 376
LENGTH: 997
TYPE: DNA
ORGANISM: Homo Sapien
US-09-902-775A-376

Query Match 6.3%; Score 52.8; DB 4; Length 997;
Best Local Similarity 56.2%; Pred. No. 1.8e-05;
Matches 99; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 665 TGCTCCATGAAGATGGGACTCGATCCACACAGCTCCCGAGTGGAGAAATGGAAT 724
DB TGTTAAAGAGGAGAGCAAGCATGTGCCACACCCGCCGCCACACGAGAAATTTGTGC 848
QY 725 TGTGTAGCCAGTGGATGCCCTAAGCATCTCCACAAATAAACCAACTCAATTATGCTT 784
DB GCTGAATCTCAAGGACTTCATAGTATTGTCTCTGATACAAATAAAATAAGTAGT 908
QY 785 AAATAAACTCAGCTGCTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
DB TTTAATGTTAAAAAATAAA 964

RESULT 13
US-09-621-976-10383
Sequence 10383, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 10383
LENGTH: 147
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-10383

Query Match 6.2%; Score 52.2; DB 4; Length 147;
Best Local Similarity 59.8%; Pred. No. 9.8e-06;
Matches 84; Conservative 2; Mismatches 55; Indels 0; Gaps 0;
QY 700 AGCTCCAGTTGGAGAAATGGAATTGTGTAGCCAAAGTGGATGCCCTAAGCATCTCCAAC 759

DB 1 AACTCTGTTTAGAGAGATCTCAATGTTTTAGTAAAGTACAAACACCATCAAAAAAAA 60
QY 760 AATAAAACAAACTCAATTATGCTTAAATAAACTCAGCTGCTTTTAAAAAATAAAAA 819
DB 61 RRAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 120
QY 820 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 840
DB 121 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 141

RESULT 14
US-08-946-026-23
Sequence 23, Application US/08946026
Patent No. 6034218
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Dillon, Davin C.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Mitcham, Jennifer L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946.026
FILING DATE: 07-OCT-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/POCKET NUMBER: 210121.424C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 375 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-946-026-23

Query Match 6.0%; Score 50.4; DB 3; Length 375;
Best Local Similarity 83.8%; Pred. No. 5.3e-05;
Matches 57; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 773 CAATTATGCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 832
DB 281 CCATTACAGCCTAAATAAAGCTTGCGCAACTTTTAAAAAATAAATAAATAAATA 340
QY 833 AAAAAAA 840
DB 341 AAAAAAA 348

RESULT 15
US-09-621-976-16320
Sequence 16320, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 22:19:21 ; Search time 587 Seconds
(without alignments)
6509.501 Million cell updates/sec

Title: US-09-937-905-1

Perfect score: 840

Sequence: 1 gaacatgtctggtgtca.....aaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2960401 seqs, 227450654 residues

Total number of hits satisfying chosen parameters: 5920802

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/PTC_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/PTCUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
 - 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
 - 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
 - 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
 - 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
 - 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
 - 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840	100.0	840	13	US-10-381-710-1
2	721.2	85.9	726	17	US-10-629-329A-3
3	598.6	71.3	1183	16	US-10-220-381-28
4	595.4	70.9	1136	13	US-10-381-710-3
5	572.2	68.1	729	17	US-10-629-329A-1
6	399.8	47.6	3784	14	US-10-044-090-252
7	378.6	45.1	402	9	US-09-933-797-133
8	279.4	33.3	384	9	US-09-796-692-4183
9	279.4	33.3	384	15	US-10-040-862-4183
10	279.4	33.3	384	16	US-10-057-475B-4183
11	279.4	33.3	384	16	US-10-154-884B-4183
12	246.4	29.3	479	10	US-09-918-995-16268
13	192.4	22.9	468	13	US-10-085-783A-34366
14	192.4	22.9	468	16	US-10-242-535A-34366

15	190.4	22.7	664	15	US-10-106-698-871	Sequence 871, App
16	150.2	17.9	473	13	US-10-085-783A-50536	Sequence 50536, A
17	150.2	17.9	473	16	US-10-242-535A-50536	Sequence 50536, A
c 18	145	17.3	532	15	US-10-029-386-2402	Sequence 2402, Ap
c 19	141.2	16.8	194	15	US-10-029-386-16102	Sequence 16102, A
20	117.2	14.0	1541	13	US-10-425-114-13586	Sequence 13586, A
21	117.2	14.0	1656	13	US-10-425-114-13377	Sequence 33377, A
22	117.2	14.0	1897	13	US-10-425-114-15898	Sequence 15898, A
23	115	13.7	1765	13	US-10-425-114-133005	Sequence 33005, A
24	113.6	13.5	792	13	US-10-424-599-14328	Sequence 14328, A
25	110.4	13.1	1830	13	US-10-425-114-12739	Sequence 12739, A
26	110.4	13.1	1886	13	US-10-424-599-130875	Sequence 130875, A
c 27	82.6	9.8	312	10	US-09-991-936-522	Sequence 522, App
c 28	82	9.7	433	9	US-09-954-456-1894	Sequence 1894, Ap
c 29	65	7.7	65	10	US-09-908-375-23950	Sequence 23950, A
c 30	64.2	7.6	255	16	US-10-131-827-8206	Sequence 8206, Ap
c 31	56.8	6.8	5884	15	US-10-311-455-529	Sequence 529, App
c 32	56.8	6.8	5884	17	US-10-257-166-39	Sequence 39, Appl
33	52.8	6.3	997	9	US-09-909-320-376	Sequence 376, App
34	52.8	6.3	997	9	US-09-909-088B-376	Sequence 376, App
35	52.8	6.3	997	9	US-09-905-291A-376	Sequence 376, App
36	52.8	6.3	997	9	US-09-902-853-376	Sequence 376, App
37	52.8	6.3	997	9	US-09-907-824-376	Sequence 376, App
38	52.8	6.3	997	9	US-09-907-841-376	Sequence 376, App
39	52.8	6.3	997	10	US-09-904-011-376	Sequence 376, App
40	52.8	6.3	997	10	US-09-906-742-376	Sequence 376, App
41	52.8	6.3	997	10	US-09-906-838-376	Sequence 376, App
42	52.8	6.3	997	10	US-09-907-613-376	Sequence 376, App
43	52.8	6.3	997	10	US-09-907-942-376	Sequence 376, App
44	52.8	6.3	997	10	US-09-904-859-376	Sequence 376, App
45	52.8	6.3	997	10	US-09-909-204-376	Sequence 376, App

ALIGNMENTS

RESULT 1

US-10-381-710-1
; Sequence 1, Application US/10381710
; Publication No. US20040052789A1
; GENERAL INFORMATION:
; APPLICANT: SHA, Shiken et al.
; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME
; FILE REFERENCE: 0230-0198P
; CURRENT APPLICATION NUMBER: US/10/381,710
; CURRENT FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Mouse macrophage cell RAW 264.7
US-10-381-710-1

Query Match 100.0%; Score 840; DB 13; Length 840;
Best Local Similarity 100.0%; Pred. No. 3.4e-210;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAACATGTCGTGCTCAAGCTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCAGG	60
Db	1	GAACCATGTCGTGCTCAAGCTCAAGGAGACTGTGCTCGCGCCGCTGTGGCGCAGG	60
QY	61	ACAAGAGACACCCCGGATTCCTGATCCCAAGAACTTTGCAACAGTTTACCAATCTGGGCT	120
Db	61	ACAAGAGACACCCCGGATTCCTGATCCCAAGAACTTTGCAACAGTTTACCAATCTGGGCT	120
QY	121	GGTCACTGGCACTGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCCTC	180
Db	121	GGTCACTGGCACTGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTCCTC	180
QY	181	CCTCAGCGCTGCAAAAGGAGCCATTCACCCAGACACATGTTTGTGTGACATTAATG	240
Db	181	CCTCAGCGCTGCAAAAGGAGCCATTCACCCAGACACATGTTTGTGTGACATTAATG	240

QY 241 AGCAGGACATAAGCGGCGCTCCAGCATCTTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300
Db 241 AGCAGGACATAAGCGGCGCTCCAGCATCTTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300
QY 301 TTTTTCATGAATGCTTATACCATGAGAGAGAGTGGCGAGTGATTCATACCATCTTAAG 360
Db 301 TTTTTCATGAATGCTTATACCATGAGAGAGAGTGGCGAGTGATTCATACCATCTTAAG 360
QY 361 CTGCTGTGATGGCTACCTTCTGTTTCCAGACAGGAGTTTAAATTTACATCAAGAGA 420
Db 361 CTGCTGTGATGGCTACCTTCTGTTTCCAGACAGGAGTTTAAATTTACATCAAGAGA 420
QY 421 TGATCAAAAGGAATTAAGGAATGTACCTCAGGAGGCTATTACAGATAGATATGTTAG 480
Db 421 TGATCAAAAGGAATTAAGGAATGTACCTCAGGAGGCTATTACAGATAGATATGTTAG 480
QY 481 TGGTACCTATTATTGAGAACACTCTCTGAGAGAGGATCTCAAGAAAGGATGCTCATG 540
Db 481 TGGTACCTATTATTGAGAACACTCTCTGAGAGAGGATCTCAAGAAAGGATGCTCATG 540
QY 541 CCATGAATGATGACCCAGACTCCTGTGCGGTTCTGTCGCGGTCATGGGCTGACGTGT 600
Db 541 CCATGAATGATGACCCAGACTCCTGTGCGGTTCTGTCGCGGTCATGGGCTGACGTGT 600
QY 601 GGGGAGAAACATGGAGAAAGCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACA 660
Db 601 GGGGAGAAACATGGAGAAAGCAAAACCATGTGTGAGTGTATGACTTACCTGTTTGACA 660
QY 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACACAGCTCCAGTTGGAGAAATG 720
Db 661 TTGCTGTCTCCATGAAGAGATGGGACTCGATCCCAACACAGCTCCAGTTGGAGAAATG 720
QY 721 GAATGTGTAGCCAGTGGATGCTTAAGCATCTCCAAATTAAGCAATCAATCAATGATG 780
Db 721 GAATGTGTAGCCAGTGGATGCTTAAGCATCTCCAAATTAAGCAATCAATCAATGATG 780
QY 781 CCTTAATATAAACTCAGTGTCTTTTAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
Db 781 CCTTAATATAAACTCAGTGTCTTTTAAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 840

RESULT 2
US-10-629-329A-3
; Sequence 3, Application US/10629329A
; Publication No. US20040086848A1
; GENERAL INFORMATION:
; APPLICANT: DARNAY, BRYANT G.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING POLYNUCLEOTIDES AND
; FILE REFERENCE: UTSC:761US
; CURRENT APPLICATION NUMBER: US/10/629,329A
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: 60/399,205
; PRIOR FILING DATE: 2002-07-29
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 726
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(726)
US-10-629-329A-3

Query Match 85.9%; Score 721.2; DB 17; Length 726;
Best Local Similarity 99.6%; Pred. No. 5.1e-179;
Matches 723; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 6 ATGCTGGCTGTCAAGCTCAAGAGAGTGTGCTCGCGCGGCTGTCGCGGAGGACAAG 65
Db 1 ATGCTGGCTGTCAAGCTCAAGAGAGTGTGCTCGCGCGGCTGTCGCGGAGGACAAG 60

QY 66 GAGCACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCTGGGTC 125
Db 61 GAGCACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCTGGGTC 120
QY 126 ACTGGCATTGGAGGGGGAATCAGTTGAGCATGGCAATGAAATCTACATTGCTCCCTCA 185
Db 121 ACCGGCATTGGAGGGGGAATCAGTTGAGCATGGCAATGAAATCTACATTGCTCCCTCA 180
QY 186 GGGCTGCAAAAGGAGCGCATTCAGCCAGAGAGACATGTTTGTGTGTGACATTAATGAGCAG 245
Db 181 GGGCTGCAAAAGGAGCGCATTCAGCCAGAGAGACATGTTTGTGTGTGACATTAATGAGCAG 240
QY 246 GACATAAGCGGGCTCCTCAGCATCTTAAGAGCTGAAAGAGCCAGTGCATCTCTTTTC 305
Db 241 GACATAAGCGGGCTCCTCAGCATCTTAAGAGCTGAAAGAGCCAGTGCATCTCTTTTC 300
QY 306 ATGAATGCTTATACCATGAGAGAGCTGGCGAGTTCATACCCACTCTAAAGCTGCT 365
Db 301 ATGAATGCTTATACCATGAGAGAGCTGGCGAGTTCATACCCACTCTAAAGCTGCT 360
QY 366 GTGATGGCTACCCCTTCTGTTCAGAGAGAGGAGTTTAAATTTACATCAAGAGATGATC 425
Db 361 GTGATGGCTACCCCTTCTGTTCAGAGAGAGGAGTTTAAATTTACATCAAGAGATGATC 420
QY 426 AAAGGAATAAGGAAATGTACTCTAGAGAGGCTATTACAGATAGATGATGTTAGTGTGA 485
Db 421 AAAGGAATAAGGAAATGTACTCTAGAGAGGCTATTACAGATAGATGATGTTAGTGTGA 480
QY 486 CCTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAAAGGAGTGGCTCATGCCATG 545
Db 481 CCTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAAAGGAGTGGCTCATGCCATG 540
QY 546 AATGAGTACCAGACTCCTGTGCGGTTCTTGTCCGGCTCATGGGGTGTAGTGTGGGA 605
Db 541 AACGAGTACCAGACTCCTGTGCGGTTCTTGTCCGGCTCATGGGGTGTAGTGTGGGA 600
QY 606 GAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGCT 665
Db 601 GAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGCT 660
QY 666 GTCTCCATGAAGAGATGGGACTCGATCCAAACAGACTCCAGTTGGAGAAATGGAATT 725
Db 661 GTCTCCATGAAGAGATGGGACTCGATCCAAACAGACTCCAGTTGGAGAAATGGAATT 720
QY 726 GTGTAA 731
Db 721 GTGTAA 726

RESULT 3
US-10-220-381-28
; Sequence 28, Application US/10220381
; Publication No. US20030207430A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BANDMAN, Olga
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; TITLE OF INVENTION: HUMAN ENZYME MOLECULES
; FILE REFERENCE: PP-0763 PCT
; CURRENT APPLICATION NUMBER: US/10/220,381
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NC 28
; LENGTH: 1183
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030207430A1 2116390CB1
US-10-220-381-28

Query Match      71.3%; Score 598.6; DB 16; Length 1183;
Best Local Similarity 84.9%; Pred. No. 1.1e-146;
Matches 694; Conservative 0; Mismatches 119; Indels 4; Gaps 2;

QY 4 CCATGTCGTGCTCAAGCTC---AAGGAGACTGTGCTCGCGCGCTGTGGCGCGCAGG 60
DB 92 CCATGTCGTGCTGTGCTCGCGGAGGAGCTGTGTTCCCGGAGATGCGGCGCGCAGG 151
QY 61 ACAAGGAGACACCCCGATTCCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 120
DB 152 ACNAGGAGATCCAGATACCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 211
QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 180
DB 212 GGGTCACTGGCACTGGAGGGGGAATTAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 271
QY 181 CCTCAGCGTGCAAAAGAGCGCATTCAGCCAGAGACATGTTGTGTGTGATTAATG 240
DB 272 CTTTCAAGAGTGCAAAAGAGCGAATTCAGCCTGAAGACATGTTGTGTGTGATTAATG 331
QY 241 AGCAGGACATAAGCGGCTCCAGCATCTAAGAGCTGAAGAAAGCCAGTGCACCTCCTC 300
DB 332 AAGAGACATAGTGACCTTCGCCATCGNAGAGCTTAAAGAAAGCCAGTGTACTCCTC 391
QY 301 TTTTCATGAATGCTTTATACATGAGAGGAGCTGGCGAGTGAATCATCCACTCTAAAG 360
DB 392 TTTTCATGAATGCTTACACATGAGAGGAGCAGGTGAGTGAATCATCCACTCTAAAG 451
QY 361 CTGCTGTGATGGCTACCTTCTGTTTCCAGCAGAGAGTGAATTAATTAACATCAAGAGA 420
DB 452 CTGCTGTGATGGCGACCTTCTTTTCCAGGAGCGGAGTTTAAATTAACATCAAGAGA 511
QY 421 TGATCAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTAG 480
DB 512 TGATAAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTAG 571
QY 481 TGGTACCTATTATTGAGAACACCTCCTGAAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
DB 572 TGGTACCTATTATTGAGAACACCTCCTGAAGAGAGGATCTCAAGAAAGGATGGCTCATG 631
QY 541 CCATGAATGATACCCAGACTCCTGCTGCGGTTCTTGTCCGCGCTCATGGGCTGATGCT 600
DB 632 CAATGAATGAATACCCAGACTCCTGCTGCGAGTCTGCTGAGTCTGAGTATATGTGT 691
QY 601 GGGGAGAACATGGGAGAGAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGACA 660
DB 692 GGGGAGAACATGGGAGAGAGCAAAACCATGTGTGAGTGTATGACTATTATTGATA 751
QY 661 TTGCTGTCTCCATGAAGAGATGGGATCGATCCCAACACAGCTCCCACTGGGAGAAATG 720
DB 752 TTGCCGTATCAATGAAGAGATGAGACTTGTATCTTCCAGCTCCCACTGGGAGAAATG 811
QY 721 GAATTTGTGAAGCAAGTGAATGCTTAAGCATCTCCCAACATTAAC - AAACCTCAATTAT 779
DB 812 GAATTTGTGAAGCAAAAGAGATGCTTAATTTATATACAGATTAAGCTTAACGTAATTAT 871
QY 780 GCCTTAATTAACCTCAGCTGCTTTTAAAGAAAGAA 816
DB 872 TATTAAATGAAGCTATTTTTTTTAAATGAATTGAAA 908

RESULT 4
US-10-381-710-3
; Sequence 3, Application US/10381710
; Publication No. US20040052789A1
; GENERAL INFORMATION:
; APPLICANT: SHA, Shiken et al.

; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME
; FILE REFERENCE: 0230-0198P
; CURRENT APPLICATION NUMBER: US/10/381,710
; CURRENT FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-381-710-3

Query Match      70.9%; Score 595.4; DB 13; Length 1136;
Best Local Similarity 84.7%; Pred. No. 7.1e-146;
Matches 692; Conservative 0; Mismatches 121; Indels 4; Gaps 2;

QY 4 CCATGTCGTGCTCAAGCTC---CAAGGAGACTGTGCTCGCGCGCTGTGGCGCGCAGG 60
DB 44 CCATGTCGTGCTGTGATGCTTGGAGGGAGACTGTGTTCCCGGAGATGCGGCGCGCAGG 103
QY 61 ACAAGGAGACACCCCGATTCCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 120
DB 104 ACAAGGAGACATCCAGATACCTGATCCCAAGACTTTTGCAAAAGTGTTCACCATCTGGGCT 163
QY 121 GGGTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 180
DB 164 GGGTCACTGGCACTGGAGGGGGAATTAGCTTGAAGCATGGCAATGGAATCTACATTTGCTC 223
QY 181 CTTCAAGCGCTGCAAAAGAGAGCGCATTCAGCCAGAGACATGTTGTGTGTGATTAATG 240
DB 224 CTTCAAGAGTGCAAAAGAGAGCGAATTCAGCCTGAAGACATGTTGTGTGTGATTAATG 293
QY 241 AGCAGGACATAAGCGGCTCCAGCATCTAAGAGCTGAAGAAAGCCAGTGTACTCCTC 300
DB 284 AAGAGACATAGTGGACCTTCGCCATCGAAGAGCTTAAAGAAAGCCAGTGTACTCCTC 343
QY 301 TTTTCATGAATGCTTTATACATGAGAGGAGCTGGCGAGTGAATCATCCACTCTAAAG 360
DB 344 TTTTCATGAATGCTTTACACATGAGAGGAGCAGGTGAGTGAATCATCCACTCTAAAG 403
QY 361 CTGCTGTGATGGCTACCTTCTGTTTCCAGGAGCAGGAGTTTAAATTAACATCAAGAGA 420
DB 404 CTGCTGTGATGGCGACCTTCTTTTCCAGGAGCGGAGTTTAAATTAACATCAAGAGA 463
QY 421 TGATCAAGGAATTAAGGAATGTACCTCAGAGGCTATTACAGATACGATGATGTAG 480
DB 464 TGATAAAGGAATTAAGGAATGTACTTCCGAGGGTATTATAGATATGATGATGTAG 523
QY 481 TGGTACTATTATTGAGAACACCTCCTGAAGAGAGGATCTCAAGAAAGGATGGCTCATG 540
DB 524 TGGTACCCATTATTGAGAAATACACCTTGGAGAGAAAGACCTCAAGAGATAGAAATGGCTCATG 583
QY 541 CCATGAATGATACCCAGACTCCTGCTGCGGTTCTTGTCCGCGCTCATGGGCTGATGCTGT 600
DB 584 CAATGAATGAATACCCAGACTCCTGCTGCGAGTACTGTGTGAGCTCATGGAGTATGTGT 643
QY 601 GGGGAGAACATGGGAGAGAAAGCAAAACCATGTGTGAGTGTATTGACTACCTGTTTGACA 660
DB 644 GGGGAGAACATGGGAGAGAAAGCAAAACCATGTGTGAGTGTATTGACTATTATTGATA 703
QY 661 TTGCTGTCTCCATGAAGAGATGGGAGTCCATCCACACAGCTCCCACTGGGAGAAATG 720
DB 704 TTGCCGTATCAATGAAGAGATGAGACTTGTATCTTCCAGCTCCCACTGGGAGAAATG 763
QY 721 GAATTTGTGAAGCAAGTGGATGCTTAAGCATCTCCCAACATTAAC - AAACCTCAATTAT 779
DB 764 GAATTTGTGAAGCAAGAGAGTCTTAATTTATATACAGATTAAGCTTAACGTAATTAT 823
QY 780 GCCTTAATTAACCTCAGCTGCTTTTAAAGAAAGAA 816
DB 824 TATTAAATGAAGCTATTTTTTTTAAATGAATTGAAA 860
```


QY 702 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCCAAGTGGATGCTTAAGCATCTCCAAACAA 761
 Db 301 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCCAAGTGGATGCTTAATATATACAGAGA 360
 QY 762 TAAACAAA 770
 Db 361 TAAAGCTAA 369

RESULT 9
 US-10-040-862-4183
 ; Sequence 4183, Application US/10040862
 ; Publication No. US20030078396A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
 ; FILE REFERENCE: 014058-013520US
 ; CURRENT APPLICATION NUMBER: US/10/040.862
 ; CURRENT FILING DATE: 2001-11-06
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: US 60/186,126
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: US 60/190,479
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: US 60/200,545
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/200,303
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/200,779
 ; PRIOR FILING DATE: 2000-05-01
 ; PRIOR APPLICATION NUMBER: US 60/200,999
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: US 60/202,084
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: US 60/206,201
 ; PRIOR FILING DATE: 2000-05-22
 ; PRIOR APPLICATION NUMBER: US 60/218,950
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: US 60/222,903
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: US 60/223,416
 ; PRIOR FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: US 60/223,378
 ; PRIOR FILING DATE: 2000-08-07
 ; PRIOR APPLICATION NUMBER: US 09/796,692
 ; PRIOR FILING DATE: 2001-03-01
 ; NUMBER OF SEQ ID NOS: 10467
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4183
 ; LENGTH: 384
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-040-862-4183

Query Match 33.3%; Score 279.4; DB 15; Length 384;
 Best Local Similarity 84.8%; Pred. No. 4.1e-63;
 Matches 313; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 402 AAAATTACATCAAGAGATGATCAAGGAATGAAGGAATGACCTCAGGAGCTATTAC 461
 Db 1 AAAATTACATCAAGAGATGATCAAGGAATGAAGGAATGATCTCCGAGGCTATTAT 60
 QY 462 AGATACATCATATGTTAGTGGTACCTATTATTGAGAACACTCCTGAAGAGAGGATCTC 521
 Db 61 AGATATGATGATGTTAGTGGTACCTATTATTGAGATACACCTGAGGAGAAAGACCTC 120
 QY 522 AAAGAAAGGATGCTCATGCCATGAATGATGATCCAGATCCCGATCTCTGTCGGTCTTGTCCGG 581

Db 121 AAAGATAGATGGCTCATGCAATGAATGAATACCCAGACTCTCTGTGTCAGTACTGTCAGA 180
 QY 582 CGTCATGGGGGTGTAOCTGTGGGGAGAAACATGGGAGAAACCAAAACCATGTCGTGAGTGT 641
 Db 181 CGTCATGGAGTATATGTCGTGGGGGAAACATGGGAGAGGCCCAAAACCATGTCGTGAGTGT 240
 QY 642 TATGACTACCTGTTTGCACATTTGCTCTCCATCAAGAAGATGGGACTCCGATCCCAACAG 701
 Db 241 TATGACTATTTTATTTGATATTTGCCGTATCAATGAAGAAAGTAGACTTGTATCTTCACAG 300
 QY 702 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCCAAGTGGATGCTTAAGCATCTCCAAACAA 761
 Db 301 CTCCAGTTGGAGAAATGGAATGTTGTAAGCCCAAGTGGATGCTTAATATATACAGAGA 360
 QY 762 TAAACAAA 770
 Db 361 TAAAGCTAA 369

RESULT 10
 US-10-057-475B-4183
 ; Sequence 4183, Application US/10057475B
 ; Publication No. US2004002068A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Clapper, Jonathan David
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Ordonez, Nadia
 ; APPLICANT: Carter, Lauren
 ; APPLICANT: McNeill, Patricia Dianne
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
 ; FILE REFERENCE: 014058-014402US
 ; CURRENT APPLICATION NUMBER: US/10/057.475B
 ; CURRENT FILING DATE: 2002-01-22
 ; PRIOR APPLICATION NUMBER: US 60/186,126
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: US 60/190,479
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: US 60/200,545
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: US 60/200,303
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/200,779
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: US 60/200,999
 ; PRIOR FILING DATE: 2000-05-01
 ; PRIOR APPLICATION NUMBER: US 60/202,084
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: US 60/206,201
 ; PRIOR FILING DATE: 2000-05-22
 ; PRIOR APPLICATION NUMBER: US 60/218,950
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: US 60/222,903
 ; PRIOR FILING DATE: 2000-08-03
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 10979
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4183
 ; LENGTH: 384
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-057-475B-4183

Query Match 33.3%; Score 279.4; DB 16; Length 384;
 Best Local Similarity 84.8%; Pred. No. 4.1e-63;
 Matches 313; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 402 AAAATTACATCAAGAGATGATCAAGGAATGAAGGAATGATCCTCAGGAGCTATTAC 461

Matches 313; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 402 AAAATTACATCAAGAGATGATCAAAAGGAATAAGAAATGTAATCTCCGGAGGGTATTAT 461

Db 1 AAAATTACATCAAGAGATGATCAAAAGGAATAAGAAATGTAATCTCCGGAGGGTATTAT 60

QY 462 AGATACGATGATATCTTGTAGTGTACCTATTATTCAGAACACTCTCTGAGAGAGGATCTC 521

Db 61 AGATATGATGATATCTTGTAGTGTACCTATTATTCAGAACACTCTCTGAGAGAGGATCTC 120

QY 522 AAAGAAAGGATGGCTCATGCGCATGAATGAGTACCAGACTCTCTGTGCGGTTCTTTGTCGG 581

Db 61 AGATATGATGATATCTTGTAGTGTACCTATTATTCAGAACACTCTCTGAGAGAGGATCTC 120

QY 582 AAAGAAAGGATGGCTCATGCGCATGAATGAGTACCAGACTCTCTGTGCGGTTCTTTGTCGG 581

Db 121 AAAGATAGATGGCTCATGCGCATGAATGAGTACCAGACTCTCTGTGAGTAGTACTGTCAGA 180

QY 582 CGTCATGGGTTGATCGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT 641

Db 181 CGTCATGGGTTGATCGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT 240

QY 642 TATGACTACCTGTTTGGACATTTGCTCTCCATGAAGAGATGGGACTCGATCCACACAG 701

Db 241 TATGACTATTTTATTTGATATTGCGGTATCAATGAAGAAAGTAGGACTTTGATCTTTCACAG 300

QY 702 CTCCAGTTGGAGAAATGGAATTTGTAAAGCCAAAGTGGATGCTTAAGCATCTCCAAACAA 761

Db 301 CTCCAGTTGGAGAAATGGAATTTGTAAAGCCAAAGTGGATGCTTAATATATACAGAGA 360

QY 762 TAAACAAA 770

Db 361 TAAAGCTAA 369

RESULT 12

US-09-918-995-16268

; Sequence 16268, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 16268

; LENGTH: 479

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)...(479)

; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-16268

Query Match 29.3%; Score 246.4; DB 10; Length 479;

Best Local Similarity 85.4%; Pred. No. 2.1e-54;

Matches 287; Conservative 0; Mismatches 46; Indels 3; Gaps 1;

QY 4 CCATGTCCTGCTGTCAAGCT---CAAGGAGACTGTTGCTCGGGCCGTTGGCGGCGAGG 60

Db 144 CCATGTCCTGCTGTGATCTTTGGGAGGAGACTGTTGTTCCCGAGATGCGGCGGCGAGG 203

QY 61 ACAAGGAGCACCCCGATTCTCTGATCCAGAACTTTGCAAAACAGTTTACCATCTGGGCT 120

Db 204 ACACGAGCATCCAGATACCTGATCCAGAACTTTGCAAAACAGTTTACCATTAGGCT 263

QY 121 GGGTCACTGGCACTGGAGGGGAATCAGCTTGAACATGGCAATGGAATCTACATGCTC 180

Db 264 GGGTCACTGGCACTGGAGGAGAAATAGCTTGAAGCATGGCGATGAAATCTACATGCTC 323

QY 181 CCTCAGGGGTGCAAAAGGAGGCGCATTCAGCCAGAGACATGTTTGTGTGACATTAATG 240

Db 1 AAAATTACATCAAGAGATGATCAAAAGGAATAAGAAATGTAATCTCCGGAGGGTATTAT 60

QY 462 AGATACGATGATATCTTGTAGTGTACCTATTATTCAGAACACTCTCTGAGAGAGGATCTC 521

Db 61 AGATATGATGATATCTTGTAGTGTACCTATTATTCAGAACACTCTCTGAGAGAGGATCTC 120

QY 522 AAAGAAAGGATGGCTCATGCGCATGAATGAGTACCAGACTCTCTGTGCGGTTCTTTGTCGG 581

Db 121 AAAGATAGATGGCTCATGCGCATGAATGAGTACCAGACTCTCTGTGAGTAGTACTGTCAGA 180

QY 582 CGTCATGGGTTGATCGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT 641

Db 181 CGTCATGGGTTGATCGTGTGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGT 240

QY 642 TATGACTACCTGTTTGGACATTTGCTCTCCATGAAGAGATGGGACTCGATCCACACAG 701

Db 241 TATGACTATTTTATTTGATATTGCGGTATCAATGAAGAAAGTAGGACTTTGATCTTTCACAG 300

QY 702 CTCCAGTTGGAGAAATGGAATTTGTAAAGCCAAAGTGGATGCTTAAGCATCTCCAAACAA 761

Db 301 CTCCAGTTGGAGAAATGGAATTTGTAAAGCCAAAGTGGATGCTTAATATATACAGAGA 360

QY 762 TAAACAAA 770

Db 361 TAAAGCTAA 369

RESULT 11

US-10-154-884B-4183

; Sequence 4183, Application US/10154884B

; Publication No. US20040005561A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; APPLICANT: Retter, Marc W.

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy

; FILE REFERENCE: 014058-013521US

; CURRENT APPLICATION NUMBER: US/10/154,884B

; CURRENT FILING DATE: 2002-05-23

; PRIOR APPLICATION NUMBER: US 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: US 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: US 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: US 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: US 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: US 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: US 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: US 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 60/222,903

; PRIOR FILING DATE: 2000-08-03

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 11290

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4183

; LENGTH: 384

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-154-884B-4183

Query Match 33.3%; Score 279.4; DB 16; Length 384;

Best Local Similarity 84.8%; Pred. No. 4.1e-63;

324	CTTCAGGAGTGC	CAAAAGGAACGAATTCAGGCCTGAGACATGTTGTTTATGATATAAATG	383
Db			
241	AGCAGGACATAGCGGGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCATCTCCTC	300	
Qy			
384	AAAAGGACATAGTGAGCCTTCGCCCTCGAGAGAGCTAAAAAAGCCAGCGTACTCCTC	443	
Db			
301	TTTTTCATGAATGCTTTATACCATGAGAGAGCTGCGC	336	
Qy			
444	TTTTTCATGAATGCTTTACACATGAGAGAGCAGGTG	479	
Db			

RESULT 13

```

US-10-085-783A-34366
; Sequence 34366, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C. C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085.783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34366
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6)..(6)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-34366

```

Query Match.	22.9%;	Score 192.4;	DB 13;	Length 468;
Best Local Similarity	78.6%;	Pred. No. 3.2e-40;		
Matches 254;	Conservative 0;	Mismatches 67;	Indels 2;	Gaps 2;

Qy	496	AGAACACTCTCGTGAAGAGAAGATCTCAAAG-AAAGGATGGCTCATGCCATGAATGAGTAC	554
Db	3	AGANCAACCTGAGGAGAAAGACCTCAAGTATAGATGGCTCATGCAATGAATGAATAC	62
Qy	555	CCGACTCCTTGTGGGGTTCCTTGTCCGGCGTCATGGGGTGTAACGTGTGGGGAGAAACATGG	614
Db	63	CCGACTCCTTGTGTCAGTACTGGTCAGACGTCATGGAGTATATGTGTGGGGGGGAAACATGG	122
Qy	615	GAGAAAGCAAAAAACATGTGTGAGCTGTATGACGTACCTGTTGTGACATTTGCTGTCTCCATG	674
Db	123	GAGAAGGCGCAAAACCAATGTGTGAGCTGTATGACTATTATTTTGATATTGCCGTATCAATG	182
Qy	675	AAGAAGATGGGACTCGATCCCAACACACAGCTCCCGATTGGAGAAAAATGGAATTTGTGTAAGCC	734
Db	183	AAGAAGATGAGCACTTTGATCCCTTCACAGCTCCCGATTGGAGAAATGGAATTTGTCTAAGCC	242
Qy	735	AAGTGGATGCGCTTAAGCATCTCCACAAATAAAAC-AAACTCAATTATGCTTTAAATAAAAC	793
Db	243	AAAAGAAAGTCTTAATTATATACAGAGATAAAGCTAAACGTAATTATTATTTAAATGGAAG	302
Qy	794	TCAGCTGCTTTTAAAAAATAAAAAA 816	
Db	303	CTATTTTTTTTAAATGCAATTGAAA 325	

RESULT 14

RESULT 14
US-10-242-535A-34366
; Sequence 34366. Application US/10242535A

```

1 Publication No. US20040013663A1
2
3 GENERAL INFORMATION:
4
5 APPLICANT: ChondroGene Inc..
6
7 APPLICANT: Liaw, C. C.
8
9 TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
10
11 FILE REFERENCE: 4231/2005
12
13 CURRENT APPLICATION NUMBER: US/10/242,535A
14
15 CURRENT FILING DATE: 2002-09-12
16
17 PRIOR APPLICATION NUMBER: US 10/085,783
18
19 PRIOR FILING DATE: 2002-02-28
20
21 PRIOR APPLICATION NUMBER: US 60/305,340
22
23 PRIOR FILING DATE: 2001-07-13
24
25 PRIOR APPLICATION NUMBER: US 60/275,017
26
27 PRIOR FILING DATE: 2001-03-12
28
29 PRIOR APPLICATION NUMBER: US 60/271,955
30
31 PRIOR FILING DATE: 2001-02-28
32
33 NUMBER OF SEQ ID NOS: 58994
34
35 SOFTWARE: PatentIn version 3.2
36
37 SEQ ID NO 34366
38
39 LENGTH: 468
40
41 TYPE: DNA
42
43 ORGANISM: Human
44
45 FEATURE:
46
47 NAME/KEY: misc feature
48
49 LOCATION: (6)..(6)
50
51 OTHER INFORMATION: n is a, c, g, or t
52
53 US-10-242-535A-34366

```

Query Match	22.9%	Score 192.4	DB 16	Length 468	
Best Local Similarity	78.6%	Pred. No. 3.2e-40			
Matches 254	Conservative	0	Mismatches 67	Indels 2	Gaps 2
QY	496	AGACACTCCTCGAAGAGAGAGATCTCAAAG- AAAGGATGGCTCATGCCATGAATAGTAGTAC	554		
DB	3	AGANCAACCTGAGGAGAAAGACCTCAAAGTAGAATGGCTCATGCAATGAATGAATAC	62		
QY	555	CCAGACTCCTGTGCGGTTCTTGTCCGCGCTCATGGGGTGTAAGTGTGGGGAGAAACATGG	614		
DB	63	CCAGACTCCTGTGCACTGACTTGGTCAGACGTCATGGAGTATATGTGTGGGGGAGAAACATGG	122		
QY	615	GAGAAAGCAAAAACCATGTGTGAGTGTATGCACTACCTGTTTGACATGTGCTGCTCCATG	674		
DB	123	GAGAAAGCCAAAACCATGTGTGAGTGTATGACTATTTATTTTGATATGCGCGTATCAATG	182		
QY	675	AAGAGATGGGACTCCGATCCAAACACACAGCTCCAGTTCGAGAAAATGGAATGTGTAAAGCC	734		
DB	183	ARGAAAGTAGGACTTTGATCCCTTCACAGCTCCAGTTCGAGTTCGAGAAAATGGAATGTGTAAAGCC	242		
QY	735	AAGTGGATGCTTAAGCATCTCCAAACAATAAAC- AAACTCAATTTATGCTTTAATAATAAAC	793		
DB	243	AAAAGAAGCTAATTTATATACAGAGATAAGCTAAACGTAATTTATTTAAATGAAAG	302		
QY	794	TCAGCTGCTTTTAAAAAATAAAAAA	816		
DB	303	CTATTTTTTTTAAATGAAATGAAA	325		

RESULT 15

US-10-106-698-871
 ; Sequence 871, Application US/10106698
 ; Publication No. US2003010960A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide;
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03

Wed Jun 2 09:13:36 2004

```

; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 871
; LENGTH: 664
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39)..(39)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-871

Query Match      22.7%; Score 190.4; DB 15; Length 664;
Best Local Similarity 79.0%; Pred. NO. 1.3e-39;
Matches 263; Conservative 0; Mismatches 67; Indels 3; Gaps 3;

QY 485 ACCATTATTGAGAACACTCTCTGAAGAGAGAGGATCTCAAGAAAGGATGGCTCATGCCAT 544
Db 1 ACCATTATTGAGAAATACACCTGAGGAGAA-GACCTCAANGATAGWATGGCTCATGCAAT 59

QY 545 GAATGAGTACCCAGACTCCTGTGCGGTTCTTGTCCGGCGTCATGGGGTGACGTGTGGG 604
Db 60 GAATGAATACCCAGACTCCTGTGCGGTTCTTGTCCGGCGTCATGGGGTGACGTGTGGG 119

QY 605 AGAAACATGGGAGAGAAACCAACCATGTGTGAGTGTATGACTACCTGTTTGACATTGC 664
Db 120 -GAAACATGGGAGAGAGCCAAACCATGTGTGAGTGTATGACTATTTTATTTGATTGC 178

QY 665 TGTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAATGGAAT 724
Db 179 CGTATCAATGAAGAAAGTAGGACTTGTATCCTTCACAGCTCCCGAGTTGGAGAAATGGAAT 238

QY 725 TGTCTAAGCCCAAGTGGATGCGTAAAGCATCTCCAAACAATAAAC-AAACTCAATTATGCTT 783
Db 239 TGTCTAAGCCCAAGTAAAGTCTAATTATATATACAGAGATAAAGCTAAACGTAATTATTATT 298

QY 784 TAAATAAACTCAGCTGCTTTTAAATAAAAAA 816
Db 299 TAAATGAAGCTATTTTAAATGAATTGAAA 331

```

Search completed: June 2, 2004, 01:15:02
Job time : 589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2004, 21:54:22 ; Search time 3630 seconds
(without alignments)
6910.257 Million cell updates/sec

Title: US-09-937-905-1

Perfect score: 840

Sequence: 1 gaacacgtctggctgtcaa.....aaaaaaaaaaaaaaaaaaaaa 840

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	92.9	897	11	AK077705 Mus muscu
2	741.2	88.2	939	13	BUS16668 AGENCOURT
3	687.2	81.8	704	13	BY756144 BY756144
4	674.2	80.3	739	14	CB595118 AGENCOURT

5	661.2	78.7	677	13	BY755126
6	641.8	76.4	645	9	AV306484
7	620	73.8	679	12	BG277254
8	618.6	73.6	642	10	AW98838
9	605.6	72.1	850	13	BUS58424
10	597	71.1	1201	9	AL525537
11	596.4	71.0	1034	13	BK362087
12	585.8	69.7	645	13	PQ193148
13	585	69.6	1124	9	AL559009
14	577	68.7	597	14	CF621082
15	574.6	68.4	710	13	BY735766
16	566.8	67.5	865	13	BUS21864
17	564.8	67.2	591	14	CD421847
18	563.8	67.1	591	10	BF608531
19	562	66.9	588	14	CF929450
20	561.6	66.9	1201	9	AL525491
21	558	66.4	770	14	CD521711
22	557.6	66.4	564	12	BG228307
23	557.2	66.3	580	10	BF019317
24	549.2	65.4	672	29	AY405038
25	546	65.0	602	9	AI172106
26	526.8	62.7	875	13	BUI82280
27	522.8	62.2	543	13	BX525605
28	517.8	61.6	600	12	BI985626
29	516	61.4	1102	9	AL580710
30	514.2	61.2	669	12	BM788265
31	510.6	60.8	714	12	BG722097
32	508.4	60.5	531	9	AL022608
33	507	60.4	572	13	BU760396
34	504.6	60.1	624	10	BB619396
35	503.6	60.0	738	12	BI185766
36	502.2	59.8	507	9	AI326904
37	490.2	58.4	552	9	AI171255
38	486.8	58.0	513	14	CD421948
39	486.4	57.9	810	12	BG209009
40	485.8	57.8	492	9	AA250391
41	485.8	57.8	923	10	BE871032
42	484.4	57.7	486	10	BE626187
43	480.8	57.2	657	12	BM762397
44	480.8	57.2	1461	10	BF970850
45	480.6	57.2	1004	12	BM450291

ALIGNMENTS

RESULT 1

AK077705

LOCUS

DEFINITION

Mus musculus 8 days embryo whole body cDNA, RIKEN full-length

enriched library, clone:5730540M11 product:monocyte macrophage 19,

full insert sequence.

AK077705

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AK077705 897 bp mRNA linear HTC 18-SEP-2003
Mus musculus 8 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:5730540M11 product:monocyte macrophage 19,
full insert sequence.

AK077705

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

TITLE

JOURNAL
MEDLINE
PUBMEDREFERENCE
AUTHORS

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE

JOURNAL
REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 897)

JOURNAL
REFERENCE
AUTHORS

Adechi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanishi, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE

JOURNAL

Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/

FEATURES

source

1. .897
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:5730540M11"
/db_xref="MGI:2393885"
/db_xref="taxon:10090"
/clone="5730540M11"
/tissue="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="8 days embryo"
115. .840
/note="unlabeled protein product; monocyte macrophage 19 (MGD) MGI:1926788, GB|NM_019735, evidence: BLASTN, 99%, match=786)
putative"
/codon_start=1
/protein_id="BAC36968.1"
/db_xref="GI:26346639"

CDS

/translation="MSGQAGDCSSRRPCGAQDKHPRFLIPELCKOFYHLNWTGTG
GSLKHGNEIYIAPSGVQKRIQPEDFVCDINEQDISGPPASKKLKKSQCCTPLFNN
AYTMRGAGAVIHTHSAAVNATILFQGEFKITHQEMIKGIRKCTGGYRYDDMLVY
PILTEPPEKDLKERMAMHANNVEYDSCAVLVRHVVVVGETWEKAKTMCEDYLED
IAYSMKMGDLDPQLPVGENV"
869. .874
/note="putative"
897
/note="putative"

polyA_signal

polyA_site

ORIGIN

Query Match 92.9%; Score 780; DB 11; Length 897;
Best Local Similarity 99.4%; Pred. No. 2.5e-115;
Matches 783; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTCTCGCGCCGTGTGGCGCGCAGG 60
DB 110 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTCTCGCGCCGTGTGGCGCGCAGG 169
QY 61 ACAAGGAGCACCCCGGATTCCTGATCCAGAACTTTGCAACAGCTTTTACCATCTGGGCT 120
DB 170 ACAAGGAGCACCCCGGATTCCTGATCCAGAACTTTGCAACAGCTTTTACCATCTGGGCT 229
QY 121 GGGTCACCTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTGCTC 180
DB 230 GGGTCACCGGCACTGGAGGGGGAATCAGCTTGAAGCATGGCAATGAATCTACATTGCTC 289
QY 181 CCTCAGGGCGTCCAAAGAGCGCATTCAGCCAGAGAGACATGTTGTGTGTGACATTAATG 240
DB 290 CCTCAGGGCGTCCAAAGAGCGCATTCAGCCAGAGAGACATGTTGTGTGTGACATTAATG 349
QY 241 AGCAGGACATAAGCGGGGCTCCAGCATCTAAGAGCTGAAAAGAGCAGTGCCTCTC 300
DB 350 AGCAGGACATAAGCGGGGCTCCAGCATCTAAGAGCTGAAAAGAGCAGTGCCTCTC 409
QY 301 TTTTCATGAATGCTTATACCATGAGAGAGCTGGCGGAGTATTCATACCCACTCTAAG 360
DB 410 TTTTCATGAATGCTTATACCATGAGAGAGCTGGCGGAGTATTCATACCCACTCTAAG 469
QY 361 CTGCTGTGATGGCTACCTCTCTGTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 420
DB 470 CTGCTGTGATGGCTACCTCTCTGTTCCAGGACAGGAGTTTAAATTTACATCAAGAGA 529
QY 421 TGATCAAGGAATAAGGAATGTACTCCTCAGGAGGCTATTACAGATCAGATGATGTTAG 480
DB 530 TGATCAAGGAATAAGGAATGTACTCCTCAGGAGGCTATTACAGATCAGATGATGTTAG 589
QY 481 TGGTACCTATTATTCAGACACTCCTGAGAGAGAGATCTCAAGAGAGAGATGGCTCATG 540
DB 590 TGGTACCTATTATTCAGACACTCCTGAGAGAGAGATCTCAAGAGAGAGATGGCTCATG 649
QY 541 CCATGAATGATACCCAGACTCCTCTGCGGTTCTTGTCCGGCGTCAATGGGGTGTAGTGT 600
DB 650 CCATGAATGATACCCAGACTCCTCTGCGGTTCTTGTCCGGCGTCAATGGGGTGTAGTGT 709
QY 601 GGGGAGAAACATGGGAGAGAGCAAAACCATGTGTGAGTGTGATGACTACCTGTTGACA 660
DB 710 GGGGAGAAACATGGGAGAGAGCAAAACCATGTGTGAGTGTGATGACTACCTGTTGACA 769
QY 661 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCCAACAGCTCCCGAGTTGGAGAAAATG 720
DB 770 TTGCTGTCTCCATGAAGAAGATGGGACTCGATCCCAACAGCTCCCGAGTTGGAGAAAATG 829
QY 721 GAATGTGTGAAGCAAGTGGATGCGCTTAAGCATCTCCCAACATTAACCAAACTCAATTATG 780
DB 830 GAATGTGTGAAGCAAGTGGATGCGCTTAAGCATCTCCCAACATTAACCAAACTCAATTATG 889
QY 781 CCTTAAT 788
DB 890 CCTTAAGT 897

RESULT 2

```

BU516668
LOCUS          939 bp      mRNA      linear      EST 12-SEP-2002
DEFINITION    AGENCOURT 10118491 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6513424 5', mRNA sequence.
ACCESSION    BU516668
VERSION      BU516668.1 GI:22824206
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
              Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 939)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue procurement: Dr. David Rowe
              cDNA Library Preparation: Invitrogen Corp
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM14087 row: b column: 17
              High quality sequence stop: 604.
FEATURES     Location/Qualifiers
              1..939
               /organism="Mus musculus"
               /mol_type="mRNA"
               /db_xref="taxon:10090"
               /clone="IMAGE:6513424"
               /tissue_type="undifferentiated limb"
               /lab_host="BHI0B (phage-resistant)"
               /clone_lib="NIH MGC 134"
               /notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
               Cloned unidirectionally. Primer: Oligo dt. Average insert
               size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
               this is a NIH_MGC Library."
ORIGIN
Query Match      88.2%; Score 741.2; DB 13; Length 939;
Best Local Similarity 97.3%; Pred. No. 3.9e-109;
Matches 786; Conservative 0; Mismatches 18; Indels 4; Gaps 3;
QY      1 GAACCATGCTGCTCTCAAGCTCAGGAGACTGTTGCTCGCGCGCTGTGGCGGCAGG 60
DB      79 GAACCATGCTGCTCTCAAGCTCAGGAGACTGTTGCTCGCGCGCTGTGGCGGCAGG 138
QY      61 ACAAGGAGACACCCCGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCT 120
DB      139 ACAAGGAGACACCCCGATTCCTGATCCAGAACTTTGCAACAGTTTACCATCTGGGCT 198
QY      121 GGCTCACTGCGCATGGAGGGGAATCAGCTTGAAGCATGCAATGAAATCAATTGCTC 180
DB      199 GGCTCACTGCGCATGGAGGGGAATCAGCTTGAAGCATGCAATGAAATCAATTGCTC 258
QY      181 CCTCAGCGTGCAGGAGCGCATTCAGCAGACAGCATGTTGTTGTGACATTAATG 240
DB      259 CCTCAGCGTGCAGGAGCGCATTCAGCAGACAGCATGTTGTTGTGACATTAATG 318
QY      241 AGCAGGACATGAAGCGGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCACCTCCTC 300
DB      319 AGCAGGACATGAAGCGGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCACCTCCTC 378
QY      301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGCGCAGTGATTCATACCCACTTAAAG 360
DB      379 TTTTCATGAATGCTTATACCATGAGAGGAGCTGCGCAGTGATTCATACCCACTTAAAG 438
QY      361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTTACATCAAGAGA 420
DB      439 CTGCTGTGATGGCTACCCCTTCTGTTCCAGACAGGAGTTTAAATTTACATCAAGAGA 498

```

FEATURES

source

RESULT 3

LOCUS

BY756144

DEFINITION

BY756144

ACCESSION

BY756144.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 704)

Oikazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I.,

Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C.,

Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,

Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,

Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,

Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,

Gariboldi, M., Giesi, C., Godzik, A., Gough, J., Grimmond, S.,

Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,

Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,

Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,

Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,

Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,

Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,

Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,

Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,

Vallada, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,

Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,

Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,

Hayatsu, N., Harozane-Kishikawa, T., Konno, H., Nakamura, M.,

Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,

Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,

Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,

Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,

Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

JOURNAL

MEDLINE

22354683

PUBMED
COMMENT

12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp;
URL: http://genome.gsc.riken.go.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
source

1. .704
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F0C0046N22"
/cell_type="1 cell"
/dev_stage="1 cell embryo"
/clone_lib="RIKEN full-length enriched, 1 cell embryo"

ORIGIN

Query Match 81.8%; Score 687.2; DB 13; Length 704;
Best Local Similarity 99.3%; Pred. No. 1.8e-100;
Matches 700; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 101 ACAGTTTACCATCTGGGTGGTCTGGCTGCTGGAGGGGGAATCAGCTTGAAGCATGG 160
DB 1 ACAGTTTACCATCTGGG-TGGGTCACGGGCACTGGAGGGGGAATCAGCTTGAAGCATGG 59

QY 161 CAATGAATCTCATTTGCTCCCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGACAT 220
DB 60 CAATGAATCTCATTTGCTCCCTCAGCGCTGCAAAAGGAGCGCATTCAGCCAGACAT 119

QY 221 GTTTGTGTGTGACATTAATGAGCAGGACATAAGCGGGCCCTCCAGCATCTAAGAAGCTGAA 280
DB 120 GTTTGTGTGTGACATTAATGAGCAGGACATAAGCGGGCCCTCCAGCATCTAAGAAGCTGAA 179

QY 281 AAAAGCCAGTGCACCTCTTTTCATGATGCTTATACATGAGGAGCTGGCCAGT 340
DB 180 AAAAGCCAGTGCACCTCTTTTCATGATGCTTATACATGAGGAGCTGGCCAGT 239

QY 341 GATTTCATACCATCTAAAGCTCTGTGTGGTCACTCCCTTCTTTTCCAGGACAGAGTT 400
DB 240 GATTTCATACCATCTAAAGCTCTGTGTGGTCACTCCCTTCTTTTCCAGGACAGAGTT 299

QY 401 TAAATATTACATCAAGAGATGATCAAGGAATAGGAATATGACCTCAGGAGGCTATTATTA 460

DB 300 TAAATATTACATCAAGAGATGATCAAGGAATAGGAATATGACCTCAGGAGGCTATTATTA 359
QY 461 CAGATAGCATGATATGTAGTGGTACCTATTATTGAGAACCACTCTGTAAGAGAAGGATCT 520
DB 360 CAGATAGCATGATATGTAGTGGTACCTATTATTGAGAACCACTCTGTAAGAGAAGGATCT 419
QY 521 CAAAGAAAGGATGGCTCATGCCATGAATGAGTACCCAGACTCCTGTGGGTTCTTGTCGG 580
DB 420 CAAAGAAAGGATGGCTCATGCCATGAACGAGTACCCAGACTCCTGTGGGTTCTTGTCGG 479
QY 591 GCCTCATGGGTGTACGTGTGGGAGAAACATGGGAGAGCAAAACCAATCTGTGAGTG 640
DB 480 GCCTCATGGGTGTACGTGTGGGAGAAACATGGGAGAGCAAAACCAATCTGTGAGTG 539
QY 641 TTATGACTTACCTGTTTGACATTTGCTCTCCATGAAGAGATGGAGCTCCGATCCAAACA 700
DB 540 TTATGACTTACCTGTTTGACATTTGCTCTCCATGAAGAGATGGAGCTCCGATCCAAACA 599
QY 701 GCTCCAGTTGGAGAAATGGAATTTGTGTAAGCAAGTGGATGCTTAAGCATCTCCAACA 760
DB 600 GCTCCAGTTGGAGAAATGGAATTTGTGTAAGCAAGTGGATGCTTAAGCATCTCCAACA 659
QY 761 ATAAACAAACTCAATTATGCTTAAATAAATACTCAGCTGCTTTT 805
DB 660 ATAAACAAACTCAATTATGCTTAAATAAATACTCAGCTGCTTTT 704

RESULT 4
CB595118
LOCUS
DEFINITION
IMAGE:30301225 5', mRNA sequence.
CB595118
CB595118.1 GI:29512974
EST.
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 739)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDCM68 row: d column: 02
High quality sequence stop: 609.
Location/Qualifiers
1. .739
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30301225"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_178"
/note="Organ: lung and heart; Vector: pDNR-LIB; Site: 1;
sfll (ggccattatggcc); Site 2: sfll (ggccgcttggcc); cDNA
made by oligo-dT priming and directionally cloned. 5' and
3' adaptors were used in cloning as follows:
5'-AAGCATGTGTCAACGAGTGGCATTCAGCCGGG-3' and
5'-ATTTCAGAGCGGCGGCGGACATG-DT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

FEATURES
source

/clone lib="RIKEN full-length enriched, submandibular gland CRL-1734 SCA-9 clone 15 cDNA"

ORIGIN	Query Match	78.7%;	Score 661.2;	DB 13;	Length 677;
	Best Local Similarity	98.7%;	Pred. No. 2.7e-96;		
	Matches 666;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;
QY	136	GAGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATTGCTCCCTCAGGCGTGCAAA	195		
Db	2	GAGCCGTAATCAGCTTGAAGCATGGCAATGAAATNTACATTGCTCCCTCAGGCGTGCAAA	61		
QY	196	AGGAGCGCATTCAGCCAGCAAGACATGTTTGTGTGTGACATTAATGAGAGGACATAAGCG	255		
Db	62	GGGAACGCATTCAGCCAGCAAGACATGTTTGTGTGTGACATTAATGAGAGGACATAAGCG	121		
QY	256	GGCCTCCAGCATCTAGAAGCTGAAAAAGCCAGTGCACCTCCCTTTTCATGAATGCTT	315		
Db	122	GGCCTCCAGCATCTAGAAGCTGAAAAAGCCAGTGCACCTCCCTTTTCATGAATGCTT	181		
QY	316	ATACCATTGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAGCTGCTGTGATGGCTA	375		
Db	182	ATACCATTGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAGCTGCTGTGATGGCTA	241		
QY	376	CCCTTCGTTTCCAGGACAGAGGTTTAAATTTACATCAAGAGATGATCAAGAGATAA	435		
Db	242	CCCTTCGTTTCCAGGACAGAGGTTTAAATTTACATCAAGAGATGATCAAGAGATAA	301		
QY	436	GGAAATGTACTCAGGAGGCTATTACAGATACGATGATCTTTAGTGGTACCTATTATTG	495		
Db	302	GGAAATGTACTCAGGAGGCTATTACAGATACGATGATCTTTAGTGGTACCTATTATTG	361		
QY	496	AGAACACTCCTGGAAGAGAGGATCTCAAGAAAGGATGGCTCATGCCATGAATGAGTACC	555		
Db	362	AGAACACTCCTGGAAGAGAGGATCTCAAGAAAGGATGGCTCATGCCATGAATGAGTACC	421		
QY	556	CAGACTCCTGTGCGGTTCTTGTCCGCGTTCATGGGGTGTACGTGTGGGAGAAACATGGG	615		
Db	422	CAGACTCCTGTGCGGTTCTTGTCCGCGTTCATGGGGTGTACGTGTGGGAGAAACATGGG	481		
QY	616	AGAAAGCAAAACCATGTGTGAGTGTTATGACTACCTGTTTGACATTGCTCTCCATGA	675		
Db	482	AGAAAGCAAAACCATGTGTGAGTGTTATGACTACCTGTTTGACATTGCTCTCCATGA	541		
QY	676	AGAAGATGGGACTCGATCCAACACAGCTCCAGTTGGAGAAAAATGGAATTTGTGAAGCCA	735		
Db	542	AGAAGATGGGACTCGATCCAACACAGCTCCAGTTGGAGAAAAATGGAATTTGTGAAGCCA	601		
QY	736	AGTGGATGCTTAGCATCTCAACAATAAAACAACTCAATATGCTTTAAATAAAATCTC	795		
Db	602	AGTGGATGCTTAGCATCTCAACAATAAAACAACTCAATATGCTTTAAATAAAATCTC	661		
QY	796	AGCTGCTTTTAAAAA	810		
Db	662	ATCTGCTTTTAAAAA	676		

RESULT 6	AV306484	645 bp	mRNA	linear	EST 24-OCT-2001
LOCUS	AV306484	RIKEN full-length enriched, 8 days embryo	Mus musculus		
DEFINITION		cDNA clone 5730540M11 3', similar to AB028863	Mus musculus mRNA		
		WMP1919, mRNA sequence.			
ACCESSION	AV306484				
VERSION	AV306484.2	GI:16393817			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					

FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

Query Match	76.4%	Score	641.8	DB 9	Length	645			
Best Local Similarity	99.7%	Pred. No.	3.4e-93						
Matches	643	Conservative	0	Mismatches	2	Indels	0	Gaps	0
QY	140	GGGAATCAGCTTGAAGCATGCGCAATGAATCTACATTCCTCCCTCAGCGCTGCAAAAGGA	199						
Db	1	GGGAATCAGCTTGAAGCATGCGCAATGAATCTACATTCCTCCCTCAGCGCTGCAAAAGGA	60						
QY	200	GGCATTTCAGCCAGAAGACATGTTTGTGTGACATTTAATGACGACGACATAAGCGGGCC	259						
Db	61	AGCATTTCAGCCAGAAGACATGTTTGTGTGACATTTAATGACGACGACATAAGCGGGCC	120						
QY	260	TCCAGCATCTAAGAAGCTGAAAAAGCCAGTGCACCTCTTTTCATGAATGCTTTATAC	319						
Db	121	TCCAGCATCTAAGAAGCTGAAAAAGCCAGTGCACCTCTCTTTTCATGAATGCTTTATAC	180						
QY	320	CATGAGAGAGCTGGCGCAGTGAATTCATACCCACTCTAAAGCTGTGTGATGCTACCT	379						
Db	181	CATGAGAGAGCTGGCGCAGTGAATTCATACCCACTCTAAAGCTGTGTGATGCTACCT	240						
QY	380	TCTGTTTCCAGGACAGAGTTTAAATTAACACATCAAGAGATGATCAAGGAATTAAGGA	439						
Db	241	TCTGTTTCCAGGACAGAGTTTAAATTAACACATCAAGAGATGATCAAGGAATTAAGGA	300						
QY	440	ATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAGTGTGATCTATTATTAGAA	499						
Db	301	ATGTACCTCAGGAGGCTATTACAGATACGATGATGTTAGTGTGATCTATTATTAGAA	360						
QY	500	CACCTCTGAAGAGAAGATCTCAAGAAAGAGTGGCTCATGCCATGAATGATGATACCCAGA	559						
Db	361	CACCTCTGAAGAGAAGATCTCAAGAAAGAGTGGCTCATGCCATGAATGATGATACCCAGA	420						
QY	560	CTCCTGTGCGGTTCTTGTCCGGCGTCATGGGTGTACGTGTGGGAGAAACATGGGAGAA	619						
Db	421	CTCCTGTGCGGTTCTTGTCCGGCGTCATGGGTGTACGTGTGGGAGAAACATGGGAGAA	480						
QY	620	AGCAAAACCATGTGTGAGTGTATGACTACTCTGTTTGACATTCGTCTCCATGAAGAA	679						
Db	481	AGCAAAACCATGTGTGAGTGTATGACTACTCTGTTTGACATTCGTCTCCATGAAGAA	540						
QY	680	GATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAAATGGAATTTGTGTAAAGCCAAATG	739						
Db	541	GATGGGACTCGATCCAAACACAGCTCCAGTTGGAGAAAATGGAATTTGTGTAAAGCCAAATG	600						
QY	740	GATGCCCTAAGCATCTCCAAACATTAACAACTCAATTTATGCTT	784						
Db	601	GATGCCCTAAGCATCTCCAAACATTAACAACTCAATTTATGCTT	645						

RESULT 7
BG277254
LOCUS
DEFINITION
BG277254.1
GI:13072373
EST.
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 679)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: ux41d07.x1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1393445
Seq primer: -40RP from Gibco
High quality sequence stop: 475.
Location/Qualifiers
1. .679

FEATURES

source
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE.3512869"
/tissue_type="maxillary process"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares NMMA maxillary process"
/vector="pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA was primed with a NotI - oligo(dT) primer [5', TGTTCCATCTCAAGTGGAGGGCGGCGGCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with NotI and cloned into the NotI and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 73.8%; Score 620; DB 12; Length 679;
Best Local Similarity 99.2%; Pred. No. 1e-89;
Matches 623; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 GAACCATGTCTGCTGTCAGCTCAAGGACATGTTGCTCGCGCGCTGTGGCGGCGAGG 60
Db 48 GAACCATGTCTGCTGTCAGCTCAAGGACATGTTGCTCGCGCGCTGTGGCGGCGAGG 107
Qy 61 ACAGGAGCACCCCGATTCCTGATCCAGAACTTTGCAAAACAGTTTACCATCTGGCT 120
Db 108 ACAGGAGCACCCCGATTCCTGATCCAGAACTTTGCAAAACAGTTTACCATCTGGCT 167
Qy 121 GGGTCACTGTCAGCTGAGGAGGAGATCAGTTCAGAACTTTGCAAAACAGTTTACCATCTGGCT 180
Db 168 GGGTCACTGTCAGCTGAGGAGGAGATCAGTTCAGAACTTTGCAAAACAGTTTACCATCTGGCT 227
Qy 181 CTTCAGCGGTGCAAAAGGAGCGCATTCAGGAGAGACATGTTTGTGTGACATTAATG 240
Db 228 CTTGAGCGGTGCAAAAGGAGCGCATTCAGGAGAGACATGTTTGTGTGACATTAATG 287
Qy 241 AGCAGGACATAGCGGCGCTCCAGCATCTAAGAGCTGAAAAGAGCCAGTGCCTCTC 300
Db 288 AGCAGGACATAGCGGCGCTCCAGCATCTAAGAGCTGAAAAGAGCCAGTGCCTCTC 347
Qy 301 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGGAGTGAATTCATCCACTCTAAAG 360
Db 348 TTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGGAGTGAATTCATCCACTCTAAAG 407
Qy 361 CTGCTGTGATGGGTACCCCTTCTGTTCCAGACAGGAGTTTAAATTAACATCAAGAGA 420
Db 408 CTGCTGTGATGGGTACCCCTTCTGTTCCAGACAGGAGTTTAAATTAACATCAAGAGA 467
Qy 421 TGATCAAGGAATAAGGAATGACCTCAGAGGCTATTACAGATAGATGATGTTAG 480
Db 468 TGATCAAGGAATAAGGAATGACCTCAGAGGCTATTACAGATAGATGATGTTAG 527
Qy 481 TGATCACTATTATTGAGAACACCTCTGAGAGAGGATCTCAAGAAAGGATGCTCATG 540
Db 528 TGATCACTATTATTGAGAACACCTCTGAGAGAGGATCTCAAGAAAGGATGCTCATG 587
Qy 541 CCATGAATGATGATCCAGACATCTGCTGCGGTTCTTGTTCGGCGCTCATGGGTTACGTT 600
Db 588 CCATGAACGAGTACCCAGACATCTCTGCGGTTCTTGTTCGGCGCTCATGGGTTACGTT 647
Qy 601 GGGGAGAAACATGGGAGAAAGCAAAAC 628
Db 648 GGGGAGAAACATGGGAGAAAGCAACACC 675

```

RESULT 8
AW988838
LOCUS      AW988838          642 bp      mRNA      linear      EST 02-JUN-2000
DEFINITION uf21b06.y1 Soares mammary_gland NLMG Mus musculus cDNA clone
IMAGE:1511987 5' similar to SW:YJZ4 YEAST P47095 HYPOTHETICAL 27.4
KD PROTEIN IN MER2-BNA1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION AW988838
VERSION   AW988838.1  GI:8184083
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 642)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          This clone is available royalty-free through LNL; contact the
          IMAGE Consortium (info@image.lnl.gov) for further information.
          MGI:938839
          Seq primer: -40RP from Gibco
          High quality sequence stop: 469.
          Location/Qualifiers
            1..642
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              /clone="IMAGE:1511987"
              /sex="female (lactating)"
              /tissue_type="mammary gland"
              /lab_host="DH10B"
              /clone_lib="Soares mammary gland NLMG"
              /note="Vector: pT73D-Pac (Pharmacia) with a modified
              polylinker; 1st strand cDNA was prepared from mammary
              gland tissue from a lactating female, and was then primed
              with a Not I - oligo(dT) primer. Double-stranded cDNA was
              ligated to Eco RI adaptors (Pharmacia), digested with Not
              I and cloned into the Not I and Eco RI sites of the
              modified pT73 vector. Library is normalized. Library
              was constructed by Bento Soares and M. Fatima Bonaudo."
ORIGIN
Query Match      73.6%; Score 618.6; DB 10; Length 642;
Best Local Similarity 99.2%; Pred. No. 1.7e-89;
Matches 621; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1  GAACCATGTCGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 60
DB      17  GAACCATGTCGGCTCAAGCTCAAGGAGACTGTTGCTCGCGCGCGTGTGGCGGCAGG 76

QY      61  ACAAGGAGCACCCCGGATTCCTGATCCCAAGACTTTTGCAACAGTTTACCACTGGGCT 120
DB      77  ACAAGGAGCACCCCGGATTCCTGATCCCAAGACTTTTGCAACAGTTTACCACTGGGCT 136

QY      121  GGGTCACTGGGCTGAGGGGGAATCAGTTGAAGCATGGCAATGAAATCACTATGCTC 180
DB      137  GGGTCACTGGGCTGAGGGGGAATCAGTTGAAGCATGGCAATGAAATCACTATGCTC 196

QY      181  CCTCAGCGCTGCAAAAGGAGCGCATTCAGCAGACACATGTTGTGTGACATTAATG 240
DB      197  CCTCAGCGCTGCAAAAGGAGCGCATTCAGCAGACACATGTTGTGTGACATTAATG 256

QY      241  AGCAGGACATAGCGGGGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCCTCCTC 300
DB      257  AGCAGGACATAGCGGGGCTCCAGCATCTAAGAGCTGAAAAAAGCCAGTGCCTCCTC 316

QY      301  TTTTCAATGAATGTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTTAAG 360
DB      317  TTTTCAATGAATGTTATACCATGAGAGGAGCTGGCGCAGTGATTCATACCCACTTAAG 376

```

```

QY      361  CTGCTGTGATGGCTACCCCTTCTGTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 420
DB      377  CTGCTGTGATGGCTACCCCTTCTGTTCCAGGACAGGAGTTTAAATTAACATCAAGAGA 436

QY      421  TGATCAAGGAATTAAGGAATGTACCTCAGAGGCTTATACAGATACGATGATGTTAG 480
DB      437  TGATCAAGGAATTAAGGAATGTACCTCAGAGGCTTATACAGATACGATGATGTTAG 496

QY      481  TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGGATCTCAAGAAAGGATGCTCATG 540
DB      497  TGGTACCTATTATTGAGAACACTCCTCGAAGAGAGGATCTCAAGAAAGGATGCTCATG 556

QY      541  CCATGAATAGTACCCAGACTCCTGTGGGTTCTGTCCGCGTCATGGGTTGACGTGT 600
DB      557  CCATGAACGAGTACCCAGACTCCTGTGGGTTCTGTCCGCGTCATGGGTTGACGTGT 616

QY      601  GGGGAGAAACATGGAGAAAGCAAAA 626
DB      617  GGGGAGAAACATGGAGAAAGCAAAA 642

RESULT 9
BU558424
LOCUS      BU558424          850 bp      mRNA      linear      EST 16-SEP-2002
DEFINITION AGENCOURT 1036389 NIH MGC 144 Mus musculus cDNA clone
IMAGE:6586918 5', mRNA sequence.
ACCESSION BU558424
VERSION   BU558424.1  GI:22908720
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 850)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: Dr. Michael Brownstein
          cDNA Library Preparation: Michael Brownstein Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/ILNL at:
          http://image.llnl.gov
          Plate: L1CM2801 row: d column: 22
          High quality sequence stop: 556.
          Location/Qualifiers
            1..850
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              /clone="IMAGE:6586918"
              /lab_host="DH10B (TI-phase-resistant)"
              /clone_lib="NIH MGC 144"
              /note="Organ: Brain; Vector: pDNR-LIB; Site 1: SfiI
              (ggccattatggc); Site 2: SfiI (ggcgcctcgcc); cDNA made
              by oligo-dT priming and directionally cloned. 5' and 3'
              adaptors were used in cloning as follows:
              5'-AAGCAGTGGTATCAACGAGATGGCCATTACGCGCGG-3' and
              5'-ATTCTAGAGCGGAGCGGCACATG-dt(30)NN-3'. Full-length
              enriched library was constructed using the Clontech
              Creator SMART kit and size-selected to contain the 0.2-0.5
              kb size fraction (other fractions present in NIH MGC 143).
              Library created in the laboratory of M. Brownstein (NIMH,
              NIH). Note: this is a NIH_MGC Library."
ORIGIN
Query Match      72.1%; Score 605.6; DB 13; Length 850;
Best Local Similarity 91.4%; Pred. No. 1.8e-87;
Matches 729; Conservative 0; Mismatches 14; Indels 55; Gaps 6;

```

JOURNAL
COMMENTUnpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12789030.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 5727.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DC012BF01Q1&cluster=5727.r. Contact:

Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID: CS0DC012BF01Q1.

Location/Qualifiers

1. .1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DC012YK02"

/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"

/clone lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source

ORIGIN

Query Match 71.1%; Score 597; DB 9; Length 1201;

Best Local Similarity 84.8%; Pred. No. 3.7e-86;

Matches 693; Conservative 0; Mismatches 120; Indels 4; Gaps 2;

```
QY 1 GAACCATGTCTGGCTCAAGCTCAAGGAGACTGTTGCTCGGGCGCTGTGGCGCAGG 60
DB 92 GAACCATGTCTGGCTCAAGCTCAAGGAGACTGTTGCTCGGGCGCTGTGGCGCAGG 151
QY 61 ACNAGGAGACCCCGATTCCTGATCCAGAACTTTGCAAAACAGTTTACCATCTGGGCT 120
DB 152 ACNAGGAGACCCCGATTCCTGATCCAGAACTTTGCAAAACAGTTTACCATCTGGGCT 211
QY 121 GGCTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGAATGAAATCTAATTGCTC 180
DB 212 GGCTCACTGGCACTGGAGGGGGAATCAGCTTGAAGCATGGAATGAAATCTAATTG 250
QY 181 CCTCAGCGGTGCAAAAGGAGCGCAATTCAGCCAGACAGACATGTTGTTGTTGACATTAATG 240
DB 251 -----GCCAAGACATGTTGTTGTTGACATTAATG 282
QY 241 AGCAGGACATAAGCGGCTCCAGCATCTAAGAGCTGMAAAAGCCAGTGCACCTCCTC 300
DB 283 AGCAGGACATAAGCGGCTCCAGCATCTAAGAGCTGMAAAAGCCAGTGCACCTCCTC 342
QY 301 TTTTCATGAATGCTTATPACATCAGAGAGAGCTGGCGCAGTGATTCATACCCTCTAAAG 360
DB 343 TTTTCATGAATGCTTATPACATCAGAGAGAGCTGGCGCAGTGATTCATACCCTCTAAAG 402
QY 361 CTGCTGTGATGGCTACCCCTCTCTGTTCCAGACAGGAGTTTAAATTTACACATCAAGAGA 420
DB 403 CTGCTGTGATGGCTACCCCTCTCTGTTCCAGACAGGAGTTTAAATTTACACATCAAGAGA 462
QY 421 TGATCAAGGAATAAGAAATGTAACCTCAGAGAGCTATTACAGATACGATGATGTTAG 480
DB 463 TGATCAAGGAATAAGAAATGTAACCTCAGAGAGCTATTACAGATACGATGATGTTAG 522
QY 481 TGGTACCTATTATTGAGACACCTCTCGAAGAGAGAGATCTCAAGAAAGAGATGCTCATG 540
DB 523 TGGTACCTATTATTGAGACACCTCTCGAAGAGAGAGATCTCAAGAAAGAGATGCTCATG 582
QY 541 CCATGAATGATACCCAGACTCTCTGTGGGTTCTGTCCGCGCTCATGGGTGTACGTGT 600
DB 583 CCATGAACGAGTACCCAGACTCTCTGTGGGTTCTGTCCGCGCTCATGGGTGTACGTGT 642
QY 601 GGGGAGAAACATGGGAG-AAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGAC 659
DB 643 GGGGAGAAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTGTTTGAC 702
QY 660 ATTGCTGTCTCCATG-AAAGAGATGGGACTCGATCCACACAGCTTCCA--GTTGGAGAA 716
DB 703 ATTGCTGTCTCCATGAAAGAAAGATGGGACTCGATCCACACAGCTTCCAAGTTGGAGAAA 762
QY 717 AATGGAATGTGT-AAGCCAAAGTGATG-CCTAAGCATCTCCAAATAAACAACCACTCA 774
DB 763 AATGGAATGTGTAAAGCCAGGTGATGCCCTAAGCATCTCCAAATAAACAACCACTCC 822
QY 775 ATTATGCTTAAATAAAA 792
DB 823 ATTATGCTTAAATAAAA 840
```

RESULT 10

AL525537 1201 bp mRNA linear EST 23-MAY-2003
LOCUS AL525537 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DC012YK02 5-PRIME, mRNA sequence.

ACCESSION AL525537.2 GI:31063401

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

REFERENCE

AUTHORS

TITLE

```

696 GGGGGAACATGGGAGAGCCAAACCATGTGTGAGTGTATGACTATTATTGATA 755
661 TTGCTGTCTCCATGAAGAAGATGGGAGCTCGATCCAAACAGAGTCCAGTGGAGAAATG 720
756 TTGCGGTATCAATGAAGAAGTAGAGTCTTGCCTTTACAGCTCCAGTGGAGAAATG 815
721 GAATTTGTGAAGCCAGTGGAGTCCCTAAGCATCTCAACAAATAAAAC-AAACTCAATTAT 779
816 GAATTTGTGAAGCCAAAGAAAGTCTTAATATATACAGATTAAGCTAAACGTATATTAT 875
780 GCCTTAAATAAAACTCAGCTGCTCTTTTAAAAAATAAAAA 816
876 TATTTAATGAAGTCTATTTTAAATGAATGAAA 912

BX362087 1034 bp mRNA linear EST 05-MAY-2003
BX362087 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CS0DJ010YP07 5-PRIME, mRNA sequence.
BX362087
BX362087.1 GI:30380601
EST.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;
1 (bases 1 to 1034)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5727.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ010CH04QPLcluster=5727.r. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue, Genoscope sequence ID : CS0DJ010CH04QPL.
FEATURES
    source
        1..1034
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DJ010YP07"
            /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
            /cell_line="JURKAT"
            /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
            10-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
    Query Match 71.0%; Score 596.4; DB 13; Length 1034;
    Best Local Similarity 84.0%; Pred. No. 4.9e-86;
    Matches 687; Conservative 8; Mismatches 119; Indels 4; Gaps 2;

QY 4 CCATGCTGGCTGTCAAGCTC---AAGGAGACTGTCTCGCGGCGGTGTGGCGGCGAGG 60
DB 94 CCATGCTGGCTGTGATGCTCGGAGGAGAGACTGTTCTCCCGAGATGCGGCGCGAGG 153
QY 61 ACAAGGACACCCCGAGTCTCTGATCCAGAACTTTCGAACAGTTTTTACCATCTGGCT 120
DB 154 ACAAGGAGCATCCAAGATACCTGTATCCAGAACTTTTGAACAGTTTTTACCATTTAGGCT 213
QY 121 GGGTCACTGGCACTGGAGGGGGAGTCACTTTGAAGCATGGCAATGAAATCTACATTGCTC 180
```

```

214 GGGTCACTGGGCACTGGAGAGGAATTAGCTTGAAGCATGGCATGAAATCTACATTGCTC 273
181 CTTCAAGGCTGCAAAAGGAGCGCATTCAGCCAGAAAGACATGTTTGTGTGTGACATTATG 240
274 CTTCAAGGAGTGCATAAGGAACGAATTCAGCCTGAAGACATGTTTGTGTGTGATATAATG 333
241 AGCAGACATAAGCCGGCTCCAGCATCTAAGAAGCTGAATAAAAGCCAGTGCACCTCCTC 300
334 AAAAGCACAAAGTGGACCTTCGCCATCGAAGAGCTAAAAAAGCCAGTGTACTCCTC 393
301 TTTTCATGAATCTTATACCATGAGAGAGCTGGGCGAGTGTATCATACCCACTCTAAAG 360
394 TTTTCATGAATCTTATACCATGAGAGAGCTGGGCGAGTGTATCATACCCACTCTAAAG 453
361 CTGCTGTGATGGCTACCTTCTTCCAGGACAGAGTTTAAATTAACATCAAGAGA 420
454 CTGCTGTGATGGCCACACTTCTTCCAGGACGGGAGTTTAAATTAACATCAAGAGA 513
421 TGATCAAGGAATAAGGAATGTACCTCAGGAGGCTTATACAGATACGATGATGTAG 480
514 TGATAAAGGAATAAAGAAATGTACTTCCGAGGGTATTATAGATATGATGTAG 573
481 TGGTACTTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGAGATGGCTCATG 540
574 TGGTACCATTATTGAGATACACTTGAGAGAAARACCTCAAAATAGAAATGGCTCATG 633
541 CCATGAATAGTACCAGACTCTCTGTGGGTTCTTGTCCGCGCTCATGGGTTGACTGT 600
634 CAATGAATGAATACCCAGACTCTCTGTGCAGTACTGTGTCTCARACCTCATGGAGTATATGT 693
601 GGGGAGAACATCGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACTCTGTTTGAACA 660
694 GGGGGAGAAACATGGAGAGAGCCAAACCATGTGTGAGTGTATGACTACTTATTTGATA 753
661 TTGCTGTCTCCATGAAGAGATGGGACTCCAGTCCAAACAGCTCCAGTTGGAGAAATG 720
754 TTGCGGTATCAAGAAGAAATAGAGACTTGATCTTCCAGCTCCAGTTGGAGAAATG 813
721 GAATTTGTGAAGCAAGTGGAGTCCATGAAGCTTCCAAACATTAAC-AAACTCAATTAT 779
814 GAATTTGTGAAGCAAGAAAGTCTAATATATACAGATTAAGCTTAAAGTAAATAT 873
780 GCCTTAATAAATCACTAGCTGCTTTTAAAAAATAAAAA 817
874 TATTAAATGAAGCTATTTTAAATGAATGAAW 911

RESULT 12
BQ193148/c
LOCUS
DEFINITION
    UI-R-DRI-clk-f-14-0-UI.s1 UI-R-DRI Rattus norvegicus cDNA clone
    UI-R-DRI-clk-f-14-0-UI 3', mRNA sequence.
ACCESSION
    BQ193148
VERSION
    BQ193148.1 GI:20368699
KEYWORDS
    Rattus norvegicus (Norway rat)
SOURCE
    Rattus norvegicus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
    Rattus.
REFERENCE
    1 (bases 1 to 645)
    AUTHORS
        Bonaldo, M.F., Lennon, G. and Soares, M.B.
    TITLE
        Normalization and subtraction: two approaches to facilitate gene
        discovery
    JOURNAL
        Genome Res. 6 (9), 791-806 (1996)
    MEDLINE
        97044477
    PUBMED
        8889548
    COMMENT
        Contact: Soares, MB
        Coordinated Laboratory for Computational Genomics
        University of Iowa
        375 Newton Road, 4156 MEERF, Iowa City, IA 52242, USA
        Tel: 319 335 8250
```


Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA Library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com). The following repetitive elements were found in this cDNA sequence: 1-22, >AT_rich#low_complexity
 Seq primer: M13 Forward
 POLVA=Yes.

FEATURES

source

Location/Qualifiers
 1..645
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DRI-clk-f-14-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-DRI"
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DRI library is a normalized Rat Osteoblast library (nREO) constructed in p317 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery, Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AAGATATCAA between the Not I cloning site and d18 stretch. The Rat Osteoblast tissue was provided by Lian & Stein of the University of Massachusetts Medical School.
 TAG TISSUE=osteoblast
 TAG LIB=UI-R-DRI
 TAG_SEQ=AAGATATCAA"

ORIGIN

Query Match 69.7%; Score 585.8; DB 13; Length 645;
 Best Local Similarity 94.3%; Pred. No. 3e-84;
 Matches 608; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 179 TCCTCAGCGGTGCAAAAGAGGCGCATTCAGCCAGAGACATGTTCTGTGTGACATTA 238
 Db 645 TCCTCAGCGGTGCAAAAGAGGCGCATTCAGCCAGAGACATGTTCTGTGTGACATTA 586

QY 239 TGAGCAGGACATAGCGGGCTCCAGCATCTAAGAGCTGAAAAAGCCAGTGCATCC 298
 Db 585 TGAGCAGGACATAGTGGGCTCCAGCATCTAAGAGCTTAAAGAGCCAGTGCATCC 526

QY 299 TCTTTTCATGAATGCTTATACCATGAGAGGAGCTGGCGGAGTGATTCATCCACTTAA 358
 Db 525 TCTTTTCATGAACCTTACACCATGAGAGGAGCGGCGGAGTGATTCATCCACTTAA 466

QY 359 AGCTGCTGTGATGCTACCTCTGTTTCCAGGACAGAGTTAAATACATCAGA 418
 Db 465 AGCTGCTGTGATGCTACCTCTGTTTCCAGGACAGAGTTAAATACATCAGA 406

QY 419 GATGATCAAGGATAGGAATGATCTCAGGAGGCTATTACAGATACGATGATATT 478
 Db 405 GATGATCAAGGATAGGAATGATCTCAGGAGGCTATTACAGATACGATGATATT 346

QY 479 AGTGTGCTATTATTGAGAACACTCTCTGAAGAGAGGATCTCAAGAAAGGATGGCTCA 538
 Db 345 AGTGTGCTATTATTGAGAACACTCTCTGAAGAGAGGACCTCAAGAAAGGATGGCTCG 286

QY 539 TGCCATGAATGATGATACCCAGACTCTGTGCGGTCTTGTCCGGCGTCAATGGGTGATCGT 598
 Db 285 TGCCATGAATGAATACCCAGACTCTGTGCGGTCTTGTCTAGTCAAGGCGTCAAGGCGGTGATCGT 226

QY 599 GTGGGGAGAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGACTACCTCTTTGA 658

Db 225 GTGGGAGAACATGGGAGAAAGCAAAACCATGTGTGAGTGTATGATTACTGTTGA 166
 QY 659 CATTGCTCTCTCCATGAAGAAGATGGGACTCGATCCAAACACAGCTCCCGAGTTGGAGAAA 718
 Db 165 CATTGCTCTCTCCATGAAGAAGATGGGCTGGATCCACGCGAGTTCCTCCAGTTGGAGACA 106
 QY 719 TGGAAATGTTGAAGCAAGTGGATGCCCTTAAGCATCTCCAAACATTAACAACAACTCAATTA 778
 Db 105 TGGAAATGTTGAAGCAAGTGGATGCCCTTAAGCATCTCCAAACATTAACAACAACTCAATTA 46
 QY 779 TGCCTTAATAAAGTCTCAGTCTCTTTTAAAAAAGAAAAAAGAAAAA 823
 Db 45 CGCCTTAATAAAGTCTCAGTCTCTTTTAAAAAAGAAAAAAGAAAAA 1

RESULT 13

AL559009

LOCUS

DEFINITION

AL559009 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED

AL559009 Homo sapiens cDNA clone CS0DJ008YC08 5-PRIME, mRNA sequence.

ACCESSION

AL559009.2 GI:31283142

VERSION

EST.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1124)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

JOURNAL

On Feb 15, 2001 this sequence version replaced gi:12904084.

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by life technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

5727.r For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DJ008BB04QP1&cluster=5727.r. Contact :

Feng Liang Email : fliang@lifetech.com/ Invitrogen Corporation 1600

http://fulllength.invitrogen.com/

Paradise Avenue Genoscope sequence ID : CS0DJ008BB04QP1.

Location/Qualifiers

1..1124

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DJ008YC08"

/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"

/cell_line="JURKAT"

/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 69.6%; Score 585; DB 9; Length 1124;
 Best Local Similarity 84.5%; Pred. No. 3.2e-84;
 Matches 692; Conservative 1; Mismatches 121; Indels 5; Gaps 3;

QY 4 CCATGTCTGGCTGCAAGCTC---AAGGAGACTGTTCTCGCGCCCTGTGGCGCGCAGG 60
 Db 138 CCATGTCTGGCTGTTGCTCGGAGGAGACTGTTGTTCCCGAGAGTGGCGCGCGCAGG 197

QY 61 ACAGGAGGACCCCGGATTCCTGATCCAGAACTTTGCAACAGTTTTTACCATCTGGCT 120
 Db 198 ACAGGAGGATCCAGATACCTGATCCAGAACTTTGCAACAGTTTTTACCATCTGGCT 257

121 GGGTCACTGGCACTGAGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTC 180
 258 GGGTCACTGGCACTGAGGGGAATGAGCTTGAAGCATGGCAATGAAATCTACATGCTC 317
 181 CTTCAAGGCTGCAAAAGAGGCGCATTCAGCCAGACATGTTTGTGTGTGACATTAATG 240
 318 CTTCAAGGCTGCAAAAGAGGCGCATTCAGCCAGACATGTTTGTGTGTGACATTAATG 377
 241 AGCAGACATAGCGGGCTCCAGCATCTAAGAGCT-GAAAAAGCGAGTGCATCCT 299
 378 AAAAGACATAGCGGGCTCCAGCATTCAGCCAGACATGTTTGTGTGTGACATTAATG 437
 300 CTTTTCATGAATGCTTATACCATGAGAGGAGTGGCGGAGTGAATTCATCCCACTTAAA 359
 438 CTTTTCATGAATGCTTATACCATGAGAGGAGTGGCGGAGTGAATTCATCCCACTTAAA 497
 360 GCTGCTGTGATGCTACCTTCTGTTTCCAGGACAGAGGTTTAAATTCACATCAAGAG 419
 498 GCTGCTGTGATGCTACCTTCTGTTTCCAGGACAGAGGTTTAAATTCACATCAAGAG 557
 420 ATGATCAAGGATTAAGGAATGCTTCTGAGGAGGCTTATACAGATACGATGATATGTTA 479
 558 ATGATCAAGGATTAAGGAATGCTTCTGAGGAGGCTTATACAGATACGATGATATGTTA 617
 480 GTGCTACCTATTATGAGACACTCTCGAAGAGAGGATCTCAAGAGAGGATGGCTCAT 539
 618 GTGCTACCTATTATGAGACACTCTCGAAGAGAGGATCTCAAGAGAGGATGGCTCAT 677
 540 GCCATGAATGAGTACCCAGACTCTGTCGCGGTTCTTGTCCGCGCTCATGGGTTGACG 599
 678 GCAATGAATGAATACCCAGACTCTGTCGAGTACTGTCGAGCGTCATGGAGTATATGTG 737
 600 TGGGGAGAACATGGGAGAGAACCAATGTCGAGTGTGATGACTACCTGTTTGC 659
 738 TGGGGAGAACATGGGAGAGAACCAATGTCGAGTGTGATGACTACCTGTTTGC 797
 660 ATTGCTGTCTCCATGAAGAGAGTGGGACTCGATCCCAACAGCTCCCGAGTTGGAGAAAT 719
 798 ATTGCTGTCTCCATGAAGAGAGTGGGACTCGATCCCAACAGCTCCCGAGTTGGAGAAAT 857
 720 GGAATGCTGTGAAGCAAGTGGGACTCGATCCCAACAGCTCCCGAGTTGGAGAAAT 778
 858 KGAATGCTGTGAAGCAAGTGGGACTCGATCCCAACAGCTCCCGAGTTGGAGAAAT 917
 779 TGCCTTAATTAATCACTGAGTGGCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG 817
 918 TTAATTAATGAGAGTGGCTTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 956

RESULT 14
 CF621082 597 bp mRNA linear EST 02-OCT-2003
 LOCUS laf22a02.y1 Gastric Epithelial Progenitor Mus musculus cDNA 5'
 DEFINITION similar to TR:Q9WV05 Q9WV05 MMRP19. ;, mRNA sequence.

CF621082
 ACCESSION CF621082.1 GI:37370011
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 597)
 REFERENCES
 AUTHORS Tidwell,R., Clifton,S., Marra,M., Hillier,L., Page,D., Martin,J., Wyllie,T., Theising,B., Bowers,Y., Gibbons,M., Ritzer,E., Bennet,J., Ronko,I., Tsagaris,I., R., Belaygorod,L., Grow,A., Maguire,L., Waterston,R. and Wilson,R.
 TITLE WashU Stem cell EST Project
 JOURNAL Unpublished (2002)

CONTACT: Jeff Gordon and Mike Lovett
 WASHU, Human Genetics Division
 Washington University School of Medicine
 COMMENT 1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified

SMART primers. The final cDNA was cloned in pAMP1 vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y.Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JI.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 585.
 Location/Qualifiers
 1..597
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /tissue_type="Gastric Epithelial Progenitor"
 /dev_stage="adult"
 /lab_host="DH5alpha"
 /clone_lib="Gastric Epithelial Progenitor"
 /note="Vector: pAMP1; This library was created from laser-captured isthmal cells from tox176 transgenic mice. 1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified SMART primers. The final cDNA was cloned in pAMP1 vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y.Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JI."

ORIGIN

Query Match 68.7%; Score 577; DB 14; Length 597;
 Best Local Similarity 98.3%; Pred. No. 7.8e-83;
 Matches 593; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 124 TCATCGGCACTGGAGGGGAAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTCCT 183
 Db 5 TAACCCCTCACTAAAGGGGGAATCAGCTTGAAGCATGGCAATGAAATCTACATGCTCCT 64
 QY -84 CAGCGGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGTGACATTAATGAGC 243
 Db 65 CAGCGGTGCAAAAGGAGCGCATTCAGCCAGAGACATGTTGTGTGTGACATTAATGAGC 124
 QY 244 AGGACATAAGCGGGCTCCAGCATCTAAGAAGCTGAAAAAGCCAGTGACCTCTCTTT 303
 Db 125 AGGACATAAGCGGGCTCCAGCATCTAAGAAGCTGAAAAAGCCAGTGACCTCTCTTT 184
 QY 304 TCATGAATGTTATACCATGAGAGAGCTGGCGCAGTGATTATACCCACTCTAAAGCTG 363
 Db 185 TCATGAATGTTATACCATGAGAGAGCTGGCGCAGTGATTATACCCACTCTAAAGCTG 244
 QY 364 CTGTGATGGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAAAATTACACATCAAGAGATGA 423
 Db 245 CTGTGATGGCTACCTCTCTGTTTCCAGGACAGGAGTTTAAAAATTACACATCAAGAGATGA 304
 QY 424 TCAAGGAATAAGGAAATGTACCTCAGGAGGCTTATACAGATACGATGATGTTAGTGG 483
 Db 305 TCAAGGAATAAGGAAATGTACCTCAGGAGGCTTATACAGATACGATGATGTTAGTGG 364
 QY 484 TACCTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAGAGAGATGGCTCATGCCA 543
 Db 365 TACCTATTATTGAGAACACTCTCTGAGAGAGAGATCTCAAGAGAGAGATGGCTCATGCCA 424
 QY 544 TGAATGATGATACCCAGACTCTCTGTCGGGCTCTGTCGGGCGCTCATGGGCTGACGTGTGGG 603
 Db 425 TGAACGATGATACCCAGACTCTCTGTCGGGCTCTGTCGGGCGCTCATGGGCTGACGTGTGGG 484
 QY 604 GAGAAACATGGGAG 663
 Db 485 GAGAAACATGGGAG 544
 QY 664 CTGTCTCCATGAGAGAGATGGGACTCGATCCCAACAGCTCCCAAGTTGGAGAA 716
 Db 545 CTGTCTCCATGAGAGAGATGGGACTCGATCCCAACAGCTCCCAAGTTGGAGAA 597

RESULT 15
 BY735766

LOCUS BY735766 710 bp mRNA linear EST 17-DEC-2002
 DEFINITION cDNA clone I0C0046N22 5', mRNA sequence.

ACCESSION BY735766
 VERSION BY735766.1 GI:27148893

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 710)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Otsato, N., Saito, R., Suzuki, H., Yamana, K., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Ciothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maitais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sadelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Iehii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Iehii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murai, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, K., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

FEATURES

source

1..710
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="I0C0046N22"
 /cell_type="1 cell"
 /dev_stage="1 cell embryo"
 /clone_lib="RIKEN full-length enriched, 1 cell embryo"

ORIGIN

Query Match 68.4%; Score 574.6; DB 13; Length 710;
 Best Local Similarity 98.3%; Pred. No. 1.8e-82;
 Matches 591; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
 QY 1 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTGTCGCGCGCGTGTGGCGCAGG 60
 DB 111 GAACCATGTCTGGCTGTCAAGCTCAAGGAGACTGTTGTCGCGCGCGTGTGGCGCAGG 170
 QY 61 ACAAGGAGACCCCGGATCTCGATCCAGAACTTTGCAACAGCTTTTACCATTCTGGCT 120
 DB 171 ACAAGGAGACCCCGGATCTCGATCCAGAACTTTTGCACACGTTTTTACCATTCTGGCT 230
 QY 121 GGGTCACCTGGCACTGGAGGGGGAATCAGTTGAAGCATGGCAATCAAAATCTACATTGCTC 180
 DB 231 GGGTCACCGGCACTGGAGGGGGAATCAGTTGAAGCATGGCAATCAAAATCTACATTGCTC 290
 QY 181 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCAGAGACATGTTTGTGTGACATTAAATG 240
 DB 291 CCTCAGGCGTGCAAAAGGAGCGCATTCAGCAGAGACATGTTTGTGTGACATTAAATG 350
 QY 241 AGCAGGACATAAGCGGGCTCAGCATCTAAGAGCTGAAAAAGCCAGTGCCTCCTC 300
 DB 351 AGCAGGACATAAGCGGGCTCAGCATCTAAGAGCTGAAAAAGCCAGTGCCTCCTC 410
 QY 301 TTTTCATCAATGCTTATACCATGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 360
 DB 411 TTTTCATCAATGCTTATACCATGAGGAGCTGGCGCAGTGATTCATACCCACTCTAAAG 470
 QY 361 CTGCTGTGATGGCTACCCCTTCTGTTCCAGGACAGAGTTTAAATACATCAAGAGA 420
 DB 471 CTGCTGTGATGGCTACCCCTTCTGTTCCAGGACAGAGTTTAAATACATCAAGAGA 530
 QY 421 TGATCAAAAGGAATAAGAAATGTACTCAGGAGGCTATTACAGATACGATGATGTAG 480
 DB 531 TGATCAAAAGGAATAAGAAATGTACTCAGGAGGCTATTACAGATACGATGATGTAG 590
 QY 481 TGGTACCTATTATTGAGAAACACTCTCTGAGAGAGAGATCTCAAGAAAGGATGGCTCATG 540
 DB 591 TGGTACCTATTATTGAGAAACACTCTGAGAGAGAGATCTC-AAGAAAGGATGGCTCATG 649
 QY 541 CCATGAATGAGTACCCAGACCTCCTGTGCGGTTCTTGTCCGCGCTCATGGGGGTACGTT 600
 DB 650 CCATGAACGAGTACCCAGACCTCCTGTGCGGTTCTTGTCCGCGCTCATGGGGGTACGTT 709
 QY 601 G 601
 DB 710 G 710

Search completed: June 2, 2004, 01:01:34
 Job time : 3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:19:15 ; Search time 60 Seconds
(without alignments)
1134.899 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQGCCRCRPGCAQDK.....SMKMKLDPDTLPVGENGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1317	100.0	241	4	AAY85635 Antigen r
2	1317	100.0	241	5	Aau77177 Murine G-
3	1317	100.0	241	5	Aau78361 Cell diff
4	1246.5	94.6	242	4	AAG67127 Amino aci
5	1239.5	94.1	242	5	Aau78360 Cell diff
6	1235.5	93.8	242	2	AAG94762 Amino aci
7	1235.5	93.8	242	4	AAY85636 Antigen r
8	1235.5	93.8	242	5	Aau77178 Human G-C
9	786	59.7	227	4	AB65485 Drosophil
10	248	18.8	129	4	AA010783 Human pol
11	202	16.9	59	4	AAG74374 Human col
12	202	15.3	212	6	ABU17451 Protein e
13	195	14.8	204	6	ABU41912 Protein e
14	172	13.1	204	6	ABU27936 Protein e
15	164	12.5	205	6	ABU15639 Protein e
16	137	10.4	181	6	ABU18884 Protein e
17	133	10.1	202	6	ABU31958 Protein e
18	131	9.9	238	3	AAY70730 Klebsiell
19	130	9.9	227	6	ABU02540 S. pneumo
20	126.5	9.6	220	6	ABU21860 Protein e
21	126.5	9.6	230	5	ABP65425 Bifidobac
22	126	9.6	230	6	ABU49674 Protein e
23	125.5	9.5	241	7	ADC94690 E. faeciu
24	125	9.5	234	6	ABU46266 Protein e
25	124.5	9.5	232	6	ABU29712 Protein e

26	124.5	9.5	233	7	ADC95935 E. faeciu
27	123.5	9.4	236	6	ABU29217 Protein e
28	123	9.3	228	6	ABU47361 Protein e
29	123	9.3	234	5	ABP27712 Streptoco
30	123	9.3	234	6	ABU46430 Protein e
31	122.5	9.3	228	6	ABU25094 Protein e
32	122.5	9.3	231	6	ABU49976 Protein e
33	121	9.2	228	6	ABU48206 Protein e
34	120	9.1	228	6	ABU28874 Protein e
35	118	9.0	242	2	AAW22376 S. pneumo
36	111.5	8.5	212	6	ABU02694 S. pneumo
37	111	8.4	238	5	ABP27711 Streptoco
38	110.5	8.4	233	6	ABU23869 Protein e
39	109.5	8.3	243	6	ABM65009 Propionib
40	108.5	8.2	75	2	AAW21952 B6-bindin
41	108.5	8.2	76	2	AAR77663 HPV B6-bi
42	103	7.8	229	2	AAW53951 Bacillus
43	102.5	7.8	210	6	ABU30461 Protein e
44	101	7.7	213	4	AAU49973 Propionib
45	101	7.7	213	6	ABM46492 Propionib

ALIGNMENTS

RESULT 1

AAU85635
ID AAY85635 standard; protein; 241 AA.

AC AAY85635;

XX 07-FEB-2001 (first entry)

DE Antigen recognised by Ab capable of inducing G-CSF activity.

XX Antigenic protein; antibody; granulocyte colony stimulating factor;

KW G-CSF; cancer therapy; bone marrow suppression; mouse.

OS Mus sp.

PN WO200060075-A1.

XX 12-OCT-2000.

PF 31-MAR-2000; 2000WO-JP002080.

XX 01-APR-1999; 99JP-00095092.

XX (NIBS) JAPAN TOBACCO INC.

XX Sha S, Aoki Y, Nishi Y;

DR WPI; 2001-024452/03.

XX N-PSDB; AAC61149.

PT Gene encoding an antigen recognizing an antibody which induces granulocyte colony stimulating factor (G-CSF) expression for gene therapy and treatment of G-CSF associated disorders e.g. the side effects of cancer therapy.

PS Claim 2; Page 49-50; 58pp; Japanese.

CC The present invention relates to a gene encoding an antigenic protein recognised by an antibody or its fragments which can induce the production of granulocyte colony stimulating factor (G-CSF). Also included in the invention are partial sequences of the gene, antibodies recognising all or part of the antigenic protein, expression vectors containing the gene and host cells transformed by the vector. The gene is used for gene therapy, and compounds identified by screening using the gene sequence are used for treatment and prevention of disorders associated with G-CSF expression such as the side effects of cancer therapy (including bone marrow suppression). The present sequence represents the murine antigen of the invention

XX SQ Sequence 241 AA;
 Query Match 100.0%; Score 1317; DB 4; Length 241;
 Best Local Similarity 100.0%; Pred. No. 4.2e-135;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSGCQAQGCCSRPCGAQDKHPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYAPS 60
 DB 1 MSGCQAQGCCSRPCGAQDKHPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYAPS 60
 QY 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTSKAA 120
 DB 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTSKAA 120
 QY 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDLKERMAMAH 180
 DB 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDLKERMAMAH 180
 QY 181 NEYPDSCAVLVRHGVVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENGI 240
 DB 181 NEYPDSCAVLVRHGVVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENGI 240
 QY 241 V 241
 DB 241 V 241

RESULT 2
 AAU771177
 ID AAU771177 standard; protein; 241 AA.
 AC AAU771177;
 DT 02-JUL-2002 (first entry)
 DE Murine G-CSF-inducible antibody binding protein, MM19.
 KW Mouse; granulocyte-colony stimulating factor; G-CSF; MM19;
 KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease;
 KW infection.
 OS Mus sp.
 PN WO200226978-A1.
 XX 04-APR-2002.
 XX 27-SEP-2001; 2001WO-JP008446.
 XX 27-SEP-2000; 2000JP-00294191.
 XX (NTSB) JAPAN TOBACCO INC.
 XX Sha S, Mukai H, Aoki Y, Nishi Y;
 XX WPI; 2002-340016/37.
 XX N-PSDB; ABK47966.
 XX Gene encoding protein binding to antibody having granulocyte-colony
 PT stimulating factor (G-CSF) inducing activity, useful for screening
 PT potential drugs treating G-CSF associated diseases.
 XX Claim 1; Page 93-94; 103pp; Japanese.
 XX The invention relates to a mouse or human gene (MM19) encoding a protein
 CC which binds to antibodies or their fragments which induce granulocyte-
 CC colony stimulating factor (G-CSF) secretion. The genes and proteins of
 CC the invention are used in diagnosis, treatment and prevention of diseases
 CC associated with G-CSF, including infections and neutrophil deficiency
 CC disease. This sequence represents a mouse G-CSF-inducible antibody
 CC binding protein, MM19

SQ Sequence 241 AA;
 Query Match 100.0%; Score 1317; DB 5; Length 241;
 Best Local Similarity 100.0%; Pred. No. 4.2e-135;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSGCQAQGCCSRPCGAQDKHPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYAPS 60
 DB 1 MSGCQAQGCCSRPCGAQDKHPRFLIPELCKQFYHLGWVTGTGGGSLKHGNEIYAPS 60
 QY 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTSKAA 120
 DB 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTSKAA 120
 QY 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDLKERMAMAH 180
 DB 121 VMTLLFPQGEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDLKERMAMAH 180
 QY 181 NEYPDSCAVLVRHGVVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENGI 240
 DB 181 NEYPDSCAVLVRHGVVWGETWEKATWCCEYDYLFDIAVSMKMGGLDPTQLPVGENGI 240
 QY 241 V 241
 DB 241 V 241

RESULT 3
 AAU78361
 ID AAU78361 standard; protein; 241 AA.
 AC AAU78361;
 DT 18-JUN-2002 (first entry)
 DE Cell differentiation stimulator associated protein #2.
 KW Cartilage cell differentiation stimulator; osteopathic;
 KW Membrane-bound transferrin-like protein; Mtf-BP; concanavalin A; ConA;
 KW membrane bound type transferrin-like protein; Mtf; cartilage disorder;
 KW bone metabolism disease; cell differentiation; cell growth;
 KW extracellular matrix related disease; mouse.
 OS Mus sp.
 PN JP2002020311-A.
 XX 23-JAN-2002.
 XX 07-JUL-2000; 2000JP-00206566.
 XX 07-JUL-2000; 2000JP-00206566.
 XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX WPI; 2002-287405/33.
 XX N-PSDB; ABK12567.
 XX A cartilage cell differentiation stimulator useful in the diagnosis of
 PT biophylaxis, cell differentiation, cell growth and construction of
 PT extracellular matrix related diseases.
 XX Claim 2; Page 9-10; 17pp; Japanese.
 XX The invention describes a cartilage cell differentiation stimulator
 CC (containing a membrane-bound transferrin-like protein (Mtf-BP) and a
 CC membrane bound type transferrin-like protein (Mtf)) and an animal-derived
 CC concanavalin-like drug. The cartilage differentiation stimulator can be
 CC used in diagnosis, prevention and treatment of cartilage and bone
 CC metabolism diseases. They can also be used for diagnosing biophylaxis,
 CC cell differentiation, cell growth and construction of extracellular
 CC matrix related diseases. Mtf-BP strongly stimulates differentiation of
 CC cartilage cells and exhibits similar action mechanism with that of plant

CC derived ConA. This is the amino acid sequence of a cartilage cell
CC differentiation stimulator associated polypeptide described in the
CC invention

XX SQ Sequence 241 AA;

Query Match 100.0%; Score 1317; DB 5; Length 241;
Best Local Similarity 100.0%; Pred. No. 4.2e-135;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSGCOAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGGGSLKHGNEIYIAPS 60
Db 1 MSGCOAQGCCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGGGSLKHGNEIYIAPS 60
QY 61 GVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTWRGAGAVIHTSKAA 120
Db 61 GVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTWRGAGAVIHTSKAA 120
QY 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGYRYDDMLVVPPIENTPEEKDKERMAHAM 180
Db 121 VMATLLFPQGEFKITHQEMIKGIRKCTSGYRYDDMLVVPPIENTPEEKDKERMAHAM 180
QY 181 NEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 240
Db 181 NEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 240
QY 241 V 241
Db 241 V 241

RESULT 4

AA067127
ID AAG67127 standard; protein; 242 AA.

XX AC AAG67127;

XX DT 13-NOV-2001 (first entry)

XX DE Amino acid sequence of a human enzyme.

XX Human; enzyme; cancer; neurological disorder; epilepsy; stroke;
KW Alzheimer's disease; Pick's disease; Huntington's disease; dementia;
KW multiple sclerosis; Parkinson's disease; amyotrophic lateral sclerosis;
KW meningitis; schizophrenia disorder; neuroskeletal disorder; allergy;
KW Addison's disease; autoimmune disease; anemia; asthma; Crohn's disease;
KW adult respiratory distress syndrome; atopic dermatitis; psoriasis;
KW diabetes mellitus; osteoporosis; pancreatitis; rheumatoid arthritis;
KW infection; genetic disorder; muscular dystrophy; Gaucher's disease;
KW Huntington's chorea; sickle cell anemia; thalassemia; atherosclerosis;
KW Von Willebrand's disease; Wilms' tumor; cell proliferative disorder;
KW leukemia; hepatitis; cirrhosis; arteriosclerosis; gene therapy.

XX OS Homo sapiens.

XX FH Location/Qualifiers

FT Modified-site 2 /note= "potential phosphorylation site"
FT Modified-site 13 /note= "potential phosphorylation site"
FT Modified-site 49 /note= "potential phosphorylation site"
FT Modified-site 57 /note= "potential phosphorylation site"
FT Modified-site 89 /note= "potential phosphorylation site"
FT Modified-site 106 /note= "potential phosphorylation site"
FT Modified-site 136 /note= "potential phosphorylation site"
FT Modified-site 148 /note= "potential phosphorylation site"
FT Modified-site 167 /note= "potential phosphorylation site"

FT Modified-site /note= "potential phosphorylation site"
FT 209 /note= "potential phosphorylation site"
FT Modified-site 216 /note= "potential phosphorylation site"
FT Modified-site 223 /note= "potential phosphorylation site"
FT Modified-site /note= "potential phosphorylation site"

WO200164896-A2.

XX 07-SEP-2001.

XX 01-MAR-2001; 2001WO-US006806.

XX 01-MAR-2000; 2000US-0186307P.

XX 28-MAR-2000; 2000US-0192532P.

XX 30-MAR-2000; 2000US-0193578P.

XX (INCY-) INCYTE GENOMICS INC.

XX Tang YT, Lu DAM, Bandman O, Yue H, Azimzai Y, Lal P, Burford N;

XX Baughn MR;

XX WPI; 2001-550184/61.

XX N-PSDB; AAH75155.

XX Novel human enzyme molecule useful for treating and preventing, e.g.,
XX cancer, genetic disorders, neurological disorders, autoimmune and
XX inflammatory disorders.

XX Claim 1; Page 117; 154pp; English.

XX The present sequence represents a human enzyme. The enzyme polynucleotide
XX and polypeptide are useful for diagnosis, treatment and prevention of
XX cancers, neurological disorders (e.g. epilepsy, stroke, Alzheimer's
XX disease, Pick's disease, Huntington's disease, dementia, multiple
XX sclerosis, Parkinson's disease, amyotrophic lateral sclerosis, bacterial
XX and viral meningitis, schizophrenic disorders (e.g. allergies, Addison's
XX disorders), autoimmune/inflammatory disorders (e.g. atopic dermatitis, diabetes mellitus,
XX anemia, asthma, Crohn's disease, atopic dermatitis, diabetes mellitus,
XX osteoporosis, pancreatitis, psoriasis, rheumatoid arthritis, and viral,
XX bacterial, fungal, parasitic, protozoal and helminthic infections),
XX genetic disorder (e.g. Duchenne and Becker muscular dystrophy, Gaucher's
XX disease, Huntington's chorea, sickle cell anemia, thalassemia, von
XX Willebrand's disease and Wilms' tumor), and cell proliferative disorder
XX (e.g. atherosclerosis, leukemia, hepatitis, cirrhosis, and
XX arteriosclerosis). The polynucleotide is also useful in somatic or
XX germ-line gene therapy

XX Sequence 242 AA;

Query Match 94.6%; Score 1246.5; DB 4; Length 242;
Best Local Similarity 94.2%; Pred. No. 2.2e-127;
Matches 228; Conservative 9; Mismatches 4; Indels 1; Gaps 1;

QY 1 MSGCOA-QGDCSRPCGAQDKHEPRFLIPELCKQFYHLGWVTGGGSLKHGNEIYIAP 59
Db 1 MSGCDAREGDCSRRCGAQDKHEPRVLIPELCKQFYHLGWVTGGGSLKHGDEIYIAP 60
QY 60 SGVQKERIQEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTWRGAGAVIHTSKA 119
Db 61 SGVQKERIQEDMFVCDINEKDISGPPSKKLKKSQCTPLFMNAYTWRGAGAVIHTSKA 120
QY 120 AVMATLLFPQGEFKITHQEMIKGIRKCTSGYRYDDMLVVPPIENTPEEKDKERMAHA 179
Db 121 AVMATLLFPQGEFKITHQEMIKGIRKCTSGYRYDDMLVVPPIENTPEEKDKERMAHA 180
QY 180 MNEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 239
Db 181 MNEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKMGGLDPTQLPVGENG 240
QY 240 IV 241

CC allowing gene inheritance to be studied through linkage analysis

XX
SQ Sequence 242 AA;
Query Match 93.8%; Score 1235.5; DB 2; Length 242;
Best Local Similarity 93.8%; Pred. No. 3.4e-126;
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;
QY 1 MSGCQA-QGDCSRPCGAQDKEHPRFLIPELCKQFVHLGWVTGTGGISLKHGNEIYIAP 59
Db 1 MSGCDWEGDCSRRCCGAQDKEHPRYLIPELCKQFVHLGWVTGTGGISLKHGDEIYIAP 60
QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSQCTPLFNNAVYMRGAGAVIHTHSA 119
Db 61 SGVQKERIQPEDMFVYDINEKDISGSPSKLKKSQCTPLFNNAVYMRGAGAVIHTHSA 120
QY 120 AVMATLLFPQBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKOLKERMHA 179
Db 121 AVMATLLFPQBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKOLKERMHA 180
QY 180 MNEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKVGLDPSQLPVGENG 239
Db 181 MNEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKVGLDPSQLPVGENG 240
QY 240 IV 241
Db 241 IV 242

RESULT 7
AAV85636
ID AAV85636 standard; protein; 242 AA.
XX AAV85636;
AC AAV85636;
DT 07-FEB-2001 (first entry)
XX Antigen recognised by Ab capable of inducing G-CSF activity.
DE Antigenic protein; antibody; granulocyte colony stimulating factor;
KW G-CSF; cancer therapy; bone marrow suppression; human.
OS Homo sapiens.
XX WO200060075-A1.
PN 12-OCT-2000.
PD 31-MAR-2000; 2000WO-JP002080.
PF 01-APR-1999; 99JP-00095092.
PR (NISB) JAPAN TOBACCO INC.
XX Sha S, Aoki Y, Nishi Y;
XX WPI; 2001-024452/03.
DR N-PSDB; AAC61150.

Gene encoding an antigen recognizing an antibody which induces granulocyte colony stimulating factor (G-CSF) expression for gene therapy and treatment of G-CSF associated disorders e.g. the side effects of cancer therapy.
Claim 3; Page 52-53; 58pp; Japanese.
The present invention relates to a gene encoding an antigenic protein recognised by an antibody or its fragments which can induce the production of granulocyte colony stimulating factor (G-CSF). Also included in the invention are partial sequences of the gene, antibodies recognising all or part of the antigenic protein, expression vectors containing the gene and host cells transformed by the vector. The gene is used for gene therapy, and compounds identified by screening using the

CC gene sequence are used for treatment and prevention of disorders associated with G-CSF expression such as the side effects of cancer therapy (including bone marrow suppression). The present sequence CC represents the human antigen of the invention

XX
SQ Sequence 242 AA;
Query Match 93.8%; Score 1235.5; DB 4; Length 242;
Best Local Similarity 93.8%; Pred. No. 3.4e-126;
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;
QY 1 MSGCQA-QGDCSRPCGAQDKEHPRFLIPELCKQFVHLGWVTGTGGISLKHGNEIYIAP 59
Db 1 MSGCDWEGDCSRRCCGAQDKEHPRYLIPELCKQFVHLGWVTGTGGISLKHGDEIYIAP 60
QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSQCTPLFNNAVYMRGAGAVIHTHSA 119
Db 61 SGVQKERIQPEDMFVYDINEKDISGSPSKLKKSQCTPLFNNAVYMRGAGAVIHTHSA 120
QY 120 AVMATLLFPQBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKOLKERMHA 179
Db 121 AVMATLLFPQBFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKOLKERMHA 180
QY 180 MNEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKVGLDPSQLPVGENG 239
Db 181 MNEYPDSCAVLVRHGVYVWGETWEKATMCCEYDYLFDIAVSMKKVGLDPSQLPVGENG 240
QY 240 IV 241
Db 241 IV 242

RESULT 8
AAU77178
ID AAU77178 standard; protein; 242 AA.
XX AAU77178;
AC AAU77178;
DT 02-JUL-2002 (first entry)
XX Human G-CSF-inducible antibody binding protein, MMRI9.
DE Human; granulocyte-colony stimulating factor; G-CSF; MMRI9;
KW antimicrobial; G-CSF-inducible antibody; neutrophil deficiency disease; infection.
XX Homo sapiens.
OS WO200226978-A1.
PN 04-APR-2002.
PD 27-SEP-2001; 2001WO-JP008446.
PF 27-SEP-2000; 2000JP-00294191.
PR (NISB) JAPAN TOBACCO INC.
XX Sha S, Mukai H, Aoki Y, Nishi Y;
XX WPI; 2002-340016/37.
DR N-PSDB; ABK47967.

Gene encoding protein binding to antibody having granulocyte-colony stimulating factor (G-CSF) inducing activity, useful for screening potential drugs treating G-CSF associated diseases.
Claim 3; Page 96-97; 103pp; Japanese.
The invention relates to a mouse or human gene (MMRI9) encoding a protein which binds to antibodies or their fragments which induce granulocyte-colony stimulating factor (G-CSF) secretion. The gene and protein of the invention are used in diagnosis, treatment and prevention of diseases

CC associated with G-CSF, including infections and neutrophil deficiency
CC disease. This sequence represents a human G-CSF-inducible antibody
CC binding protein, MMRL9
XX
SQ Sequence 242 AA;

Query Match 93.8%; Score 1235.5; DB 5; Length 242;
Best Local Similarity 93.8%; Pred. No. 3.4e-126;
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MSGCQA-QGDCSPCAQDKEHPRFLPELCKQFYHLGWVTGGGSLKHGNEIYIAP 59
Db 1 MSGCDAWEGDCSRCAQDKEHPRYLPELCKQFYHLGWVTGGGSLKHGDEIYIAP 60
Qy 60 SGVOKERIQEDMFVCDINEODISPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTSKA 119
Db 61 SGVOKERIQEDMFVCDINEODISPPSKKLKKSQCTPLFMNAYTMRGAGAVIHTSKA 120
Qy 120 AVMATLFPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
Db 121 AVMATLFPFGREFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180
Qy 180 MNEVPDSCAVLRRHGVVWGETWEKAKTMCCEYDYLFDIAVSMKKGLDPTOLPVGNG 239
Db 181 MNEVPDSCAVLRRHGVVWGETWEKAKTMCCEYDYLFDIAVSMKKGLDPSQLPVGNG 240
Qy 240 IV 241
Db 241 IV 242

RESULT 9
ABB65485
ID ABB65485 standard; protein; 227 AA.
XX
AC ABB65485;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 23247.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL09588.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 23247; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 227 AA;

Query Match 59.7%; Score 786; DB 4; Length 227;
Best Local Similarity 70.5%; Pred. No. 4.3e-77;
Matches 148; Conservative 20; Mismatches 42; Indels 0; Gaps 0;

Qy 21 EHRFLIPELCKQFYHLGWVTGGGSLKHGNEIYIAPSGVOKERIQEDMFVCDINEQ 80
Db 22 EHRFLIPELCKQFYHLGWVTGGGSLKHGNEIYIAPSGVOKERIQEDMFVCDITGK 71
Qy 81 DISPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTSKAAVMAATLFPFGQEFKITHQEMI 140
Db 72 DLQLPPEIKGLKKSQCTPLFMNAVQHRQAGAVIHTSHQAVMATLWFGKTRCTHLEMI 131
Qy 141 KGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHANEYDSCAVLRRHGVVWNG 200
Db 132 KGVYDEADKRLRYDEELVPIIENTPFERDLADSMYAAAMEYPCSAILVRRHGVVWNG 191
Qy 201 ETWEKAKTMCCEYDYLFDIAVSMKKGLDP 230
Db 192 QNWEKAKTMCCEYDYLFDIAVSMKKGLDP 221

RESULT 10
AAO10783
ID AAO10783 standard; protein; 129 AA.
XX
AC AAO10783;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 24675.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI90714.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 24675; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,

e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Sequence 129 AA;

Query Match 18.8%; Score 248; DB 4; Length 129;
Best Local Similarity 46.4%; Pred. No. 1.2e-18;
Matches 51; Conservative 4; Mismatches 7; Indels 48; Gaps 2;

QY 1 MSGCQA-QGDCSRPGQAQ-----18
DB 20 MSGCNARKGDCSRGSHLXNXPDTDLNFFLTSAKVKCFKKEIYSQTVVXSPG 79

QY 19 -----DKHPRFLIPCLCKQFVHLGWVTGGGSLKKGHNEIYIAPSGV 62
DB 80 KXWXTXDKHPRVLIPLXLCIQFVHLGWVTGGGIIILKHGEIYIAPSGV 129

RESULT 11
AAG74374
ID AAG74374 standard; protein; 59 AA.
XX AAG74374;
XX
DT 03-SEP-2001 (first entry)
DE Human colon cancer antigen protein SEQ ID NO:5138.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX Homo sapiens.
XX
XX W0200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US026524.
XX
XX 29-SEP-1999; 99US-0157137P.
XX
XX 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
XX
XX N-PSDB; AAH33805.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 11; Page 6841; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77798 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patients own production of P. Additionally, N may be used
XX to produce the colon cancer-associated P, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAH77789 represent

sequences used in the exemplification of the present invention. N.B.
Pages 666 to 682 and page 7053 of the sequence listing were missing at
time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922

SQ Sequence 59 AA;

Query Match 16.9%; Score 222; DB 4; Length 59;
Best Local Similarity 95.2%; Pred. No. 2.8e-16;
Matches 40; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 203 GETWEKAKTMCECYDYLFDIAVSMKMGDLPTQLPVGENGIV 241
DB 18 GETWEKAKTMCECYDYLFDIAVSMKMGDLPTQLPVGENGIV 59

RESULT 12
ABU17451
ID ABU17451 standard; protein; 212 AA.
XX
XX ABU17451;
XX
DT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #2978.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Bacillus anthracis.
XX
XX W0200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX
XX Wali D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA21321.

New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 45375; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 212 AA;
Query Match 15.3%; Score 201; DB 6; Length 212;
Best Local Similarity 28.8%; Pred. No. 3.5e-13;
Matches 60; Conservative 29; Mismatches 93; Indels 26; Gaps 6;
QY 29 ELCKQFYHL-----GWTGTGGGSLKHGNE---IYAPSGVQKERIQPEDMFVC 75
Db 3 QLFQWYDLSEIKKELTTRNWPATSGNISIXVSHPELTLITAGKDKTKTTPDDFLIV 62
QY 76 DINEQDISGPPASKLLKKSQCTPLFNAYTMRGAGAVIHTSKAAVMATLLPFGQEFKI 135
Db 63 D-----HLGVFVLETELRPSAETILHTHIYNNWAGCVLHVHTDNNVITNLY-SDAVTLQ 117
QY 136 HOEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHMANEYDPSCAVLVRRHG 195
Db 118 NQEIILKALDIWEGA-----TTHIPIENHAHPTLGENFRKHQ--GDSGAVLIRNHG 169
QY 196 VYVWGTEWKAKTMCBCYDLFDIAVM 223
Db 170 ITWVGDRSFDKAKRLEAYEPLFQFIKL 197
RESULT 13
ABU41912
ID ABU41912 standard; protein; 204 AA.
XX
AC ABU41912;
XX
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #27439.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Pseudomonas syringae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA45782.
XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 69836; 1766bp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 204 AA;
Query Match 14.8%; Score 195; DB 6; Length 204;
Best Local Similarity 25.2%; Pred. No. 1.5e-12;
Matches 53; Conservative 36; Mismatches 93; Indels 28; Gaps 6;
QY 21 EHRFLIPELCKQFYHLGWTGTGGGSLK-HGNEIYIAPSGVQKERIQPEDMFVCDI-- 77
Db 12 EAGRFL-----YCRGWSPATSSNYSVRLSASEALLTVSGHKHKGQGLGDDVLATDLA 63
QY 78 NEQDISGPPASKLLKKSQ---CTPLFNAYTMRGAGAVIHTSKAAVMATLLPFGQEFKI 134
Db 64 NSLEPKKPSAETLLHTLYLCLRP-----QVGAVLHTHSVNTVLSRLTASDLVLF 114
QY 135 THOEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHMANEYDPSCAVLVRRH 194
Db 115 EDVELQKAF-----NGVLTHESQVWVPIFDNDQDIARLANVQFPLDHAPECAGYLIRGH 169
QY 195 GVTVWGTEWKAKTMCBCYDLFDIAVMK 224
Db 170 GLYTWGARMSDALRQIEAFEFLECELMKR 199
RESULT 14
ABU27936
ID ABU27936 standard; protein; 204 AA.
XX
XX ABU27936;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #13463.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Enterobacter cloacae.

Wed Jun 2 09:13:36 2004

us-09-937-905-2.rag

CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*, the present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: the sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 205 AA;

Query Match 12.5%; Score 164; DB 6; Length 205;
Best Local Similarity 25.9%; Pred. No. 3.7e-09;
Matches 51; Conservative 28; Mismatches 98; Indels 20; Gaps 5;
QY 24 RFLPELCKQFYHLGWTCGGSISLKHGNE-IYIAPSGVQKRIQPEDMFVCDINEQDI 82
DB 17 RFL-----YGRGWSPATSSNYGARDLDEQALITVSGKHGQGFDDVLA-----TDL 63
QY 83 SGPPASKKLKKSQCTPLFMNAYTMRGA-GAVIHTHSKAAMVATLLFPQGEFKITHQEMIK 141
DB 64 AGNSLEPGKKPSAETLLHTQLYANNPAIGAVLHTHSVNATVLSRLVRGDLVLQDYELQK 123
QY 142 GIRKCTSGGYRYDDMLVVPILIENTPEEKOLKERMAMAHMNEYDPDSCAVLVRHGVYVWGE 201
DB 124 AF-----AGVTTHEGQVEVEIFDNDQDIARLASRVQPWLEAHPHCFGYLIRGHGYTNGA 178
QY 202 TWEKATMCECYDLYFD 218
DB 179 RMSDALRQVEAFELFE 195

Search completed: June 1, 2004, 13:33:22
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:32:21 ; Search time 22 seconds
(without alignments)
565.539 Million cell updates/sec

Title: US-09-937-905-2
Perfect score: 1317
Sequence: 1 MSCGQAGDCSPPCGAQDK.....SMKKMGIDLPQLPVGENGIV 241

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pdp.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pdp.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pdp.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pdp.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pdp.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	164	12.5	238	4	US-09-252-991A-30160	Sequence 30160, A
2	142	10.8	205	4	US-09-489-039A-13421	Sequence 13421, A
3	139	10.6	231	4	US-09-172-952-32	Sequence 32, Appl
4	131	9.9	238	4	US-09-172-952-18	Sequence 18, Appl
5	128.5	9.8	231	4	US-09-172-952-33	Sequence 33, Appl
6	125.5	9.5	241	4	US-09-107-532A-4317	Sequence 4317, Ap
7	124.5	9.5	233	4	US-09-107-532A-5562	Sequence 5562, Ap
8	123.5	9.4	241	4	US-09-134-000C-6065	Sequence 6065, Ap
9	118	9.0	242	3	US-08-472-534-2	Sequence 21, Appl
10	113.5	8.6	240	3	US-08-826-842B-21	Sequence 10152, A
11	109.5	8.3	225	4	US-09-489-039A-10152	Sequence 20, Appl
12	109.5	8.3	231	3	US-08-926-842B-20	Sequence 11, Appl
13	108.5	8.2	75	2	US-08-840-683-11	Sequence 11, Appl
14	108.5	8.2	75	2	US-08-555-722-11	Sequence 11, Appl
15	108.5	8.2	75	3	US-09-384-301-11	Sequence 12402, A
16	106	8.0	244	4	US-09-489-039A-8943	Sequence 8943, Ap
17	105.5	8.0	285	4	US-09-489-039A-12402	Sequence 14, Appl
18	105	8.0	229	3	US-08-926-842B-14	Sequence 9000, Ap
19	101.5	7.7	216	4	US-09-489-039A-9000	Sequence 11424, A
20	89.5	6.8	225	4	US-09-489-039A-11424	Sequence 13768, A
21	89.5	6.8	225	4	US-09-489-039A-13768	Sequence 15, Appl
22	85	6.5	462	4	US-09-129-112-15	Sequence 2, Appl
23	82.5	6.3	1729	4	US-09-553-690-2	Sequence 10, Appl
24	81.5	6.2	260	3	US-08-081-929-10	Sequence 8023, Ap
25	80.5	6.1	399	4	US-09-489-039A-8023	Sequence 2, Appl
26	80	6.1	234	1	US-08-850-118-2	Sequence 2, Appl
27	80	6.1	234	2	US-09-008-253-2	Sequence 2, Appl

28	80	6.1	234	3	US-09-093-335-2	Sequence 2, Appl
29	80	6.1	536	4	US-08-426-509A-12	Sequence 12, Appl
30	80	6.1	536	4	US-08-232-545-12	Sequence 12, Appl
31	80	6.1	536	5	PCT-US95-05008-12	Sequence 12, Appl
32	79	6.0	381	3	US-08-857-076-106	Sequence 106, Appl
33	79	6.0	558	4	US-09-252-991A-16908	Sequence 16908, A
34	79	6.0	1724	3	US-08-857-076-12	Sequence 12, Appl
35	78	5.9	544	4	US-09-614-912-72	Sequence 72, Appl
36	77.5	5.9	298	4	US-09-540-236-2717	Sequence 2717, Ap
37	76	5.8	537	4	US-08-426-509A-11	Sequence 11, Appl
38	76	5.8	537	4	US-08-232-545-11	Sequence 11, Appl
39	76	5.8	537	5	PCT-US95-05008-11	Sequence 11, Appl
40	76	5.8	567	3	US-08-841-483-2	Sequence 2, Appl
41	76	5.8	567	3	US-09-382-911-2	Sequence 2, Appl
42	75.5	5.7	307	2	US-08-216-894-6	Sequence 6, Appl
43	75.5	5.7	307	3	US-09-115-746-6	Sequence 6, Appl
44	75.5	5.7	502	3	US-08-123-934A-4	Sequence 4, Appl
45	75.5	5.7	502	4	US-09-874-628-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-252-991A-30160
; Sequence 30160, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30160
; LENGTH: 238
; TYPE: PPT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30160

Query Match	12.5%	Score 164;	DB 4;	Length 238;
Best Local Similarity	25.9%	Pred. No. 9.9e-10;		
Matches	51;	Conservative	28;	Mismatches 98; Indels 20; Gaps 5;
QY	24	RFLPELCKQFHLGWTGTGGISLKHGNE-IYIAPSGVQKRIQPEPMFVCDINEQDI	82	
DB	50	RFL-----YGRGWSPATSNYSARLDEQALITVSGKHGQGLGFDVLA-----TDL	96	
QY	83	SGPPASKKLKKSQCTPLFMNAYTMGA-GAVITHSKAAVMATLLPPGQEFKITHQEMTK	141	
DB	97	AGNSLEPGKPAEFLITLQYVAMPALGAVLTHSVNATVLSRLVRGRLVQDELQK	156	
QY	142	GIRKCTSGGYRDDMLVVPPIENTPEEKLMKMAHANEYDSCAVLVRHGVTVNGE	201	
DB	157	AF-----AGVTHGEGVEVPIFDNDQDIARLASRVQVLEAHPHCPGYLIRHGLYTWGA	211	
QY	202	TWEAKTMCEVDYLPD	218	
DB	212	RMSDALRQVEAEFLFE	228	

RESULT 2
US-09-489-039A-13421
; Sequence 13421, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Berton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: 2709.2004001
;; CURRENT APPLICATION NUMBER: US/09/489, 039A
;; CURRENT FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: US 60/117,747
;; PRIOR FILING DATE: 1999-01-29
;; NUMBER OF SEQ ID NOS: 14342
;; SEQ ID NO 13421
;; LENGTH: 205
;; TYPE: PRT
;; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13421

Query Match 10.8%; Score 142; DB 4; Length 205;
Best Local Similarity 23.7%; Pred. No. 2e-07;
Matches 45; Conservative 37; Mismatches 96; Indels 12; Gaps 4;
QY 38 GWTGTGGGISLKHGNE--IYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLKKSQC 96
DB 23 GWAPATGNNMSVRQDDTWCLWLSSEGRDKSLTTEDEL-----QVEIATNQAPSGRKPSAE 77
QY 97 TPLFMNAYTM-RGAGAVIHTHSAAVWATLLFPQGEFKITHQEMIKGIRKCTSGGYRYD 155
DB 78 TGLHTLVYRLFPPEANVVLHVHTVNAVLSRIKSDTLAQGYEMQK-----TILSGQSHL 132
QY 156 DMLVVPPIENTPEKDKERMAHANNEYDPDSCAVLVRHGVVYVWGETWEKAKTMCECYD 215
DB 133 DTVPALFDNDQDIDALAAIADYAOQTRPLRYGLLRHGLTGWKDIQARQLEGLEF 192
QY 216 LFDIAVSMKK 225
DB 193 LFECELMRRR 202

RESULT 3

US-09-172-952-32
;; Sequence 32, Application US/09172952
;; Patent No. 6368793
;; GENERAL INFORMATION:
;; APPLICANT: Hoch, James
;; APPLICANT: Dartois, Veronique
;; TITLE OF INVENTION: METABOLIC SELECTION METHODS
;; FILE REFERENCE: 234/191
;; CURRENT APPLICATION NUMBER: US/09/172,952
;; CURRENT FILING DATE: 1998-10-14
;; NUMBER OF SEQ ID NOS: 33
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 32
;; LENGTH: 231
;; TYPE: PRT
;; ORGANISM: Yias-EC
US-09-172-952-32

Query Match 10.6%; Score 139; DB 4; Length 231;
Best Local Similarity 27.8%; Pred. No. 5.1e-07;
Matches 65; Conservative 27; Mismatches 96; Indels 46; Gaps 13;
QY 26 LIPELCKOFYHLGWGTGGGISLKHGNE--IYIAPSGVQKRIQPEDMFVCDINEQDIS 83
DB 10 LAANLALPAHL--VTFWGNVSAVDETRWNVKPSGVYDVMTADDMVVVEI-----AS 63
QY 84 GPPASKLKKSQCTPLFMNAYTMRG-AGAVIHTHSAKAVM--ATLLFPQGEFKITHQEM 139
DB 64 GKVEGSKXPSSDTPTHALYRYAEIGGIVHTSHRATIWSQAGLDLPA--WGTHADY 121
QY 140 IKGIRKCTS-----GGYRYDDMLVVPPIENTPEKDKERMAHANNEYDPDSCAVLVR 192
DB 122 FYGALPCTQMTAEELINGEYEQ---TGEVILTEFER-----GRSPAQIP---AVLVH 169
QY 193 RHGVVYWG---ETWEKAKTMCEC-YDYLFDIAVSMKKGLDPTQLPVGENGIV 241
DB 170 SHGPPAWGNKAADAVHNAVLEECATMGLFSRQLA-----POLPAMQNEL 215

RESULT 4

US-09-172-952-18
;; Sequence 18, Application US/09172952
;; Patent No. 6368793
;; GENERAL INFORMATION:
;; APPLICANT: Hoch, James
;; APPLICANT: Dartois, Veronique
;; TITLE OF INVENTION: METABOLIC SELECTION METHODS
;; FILE REFERENCE: 234/191
;; CURRENT APPLICATION NUMBER: US/09/172,952
;; CURRENT FILING DATE: 1998-10-14
;; NUMBER OF SEQ ID NOS: 33
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 18
;; LENGTH: 238
;; TYPE: PRT
;; ORGANISM: Yias-Ko
US-09-172-952-18

Query Match 9.9%; Score 131; DB 4; Length 238;
Best Local Similarity 28.0%; Pred. No. 3.9e-06;
Matches 60; Conservative 21; Mismatches 87; Indels 46; Gaps 11;
QY 38 GWTGTGGGISLKHGNE--IYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLKKSQ 95
DB 20 GLVTFWGNVSAVDETRWNVKPSGVYEVMTADDMVVVEM-----ASGVVEGKKPSS 75
QY 96 CTLEFNAY--TMRGAGAVIHTHSAKAVM--ATLLFPQGEFKITHQEMIKGIRKCTS--- 148
DB 76 DTPTHALYRYDQIGGIVHTSHRATIWSQAGLDLPA--WGTHADYFYGAIPCTRRMT 133
QY 149 ---GGYRYDDMLVVPPIENTPEKDKERMAHANNEYDPDSCAVLVRHGVVYWG----- 200
DB 134 VEEINGEYEQ---TGEVILKTEQRLDPA-----QIP---AVLVHSHGPFANGKDA 181
QY 201 ETWEKAKTMCECYDYLFDIAVSMKKGLDPTQLP 234
DB 182 DAVHNAVLEEC-----AYMGLFSRQMP 204

RESULT 5

US-09-172-952-33
;; Sequence 33, Application US/09172952
;; Patent No. 6368793
;; GENERAL INFORMATION:
;; APPLICANT: Hoch, James
;; APPLICANT: Dartois, Veronique
;; TITLE OF INVENTION: METABOLIC SELECTION METHODS
;; FILE REFERENCE: 234/191
;; CURRENT APPLICATION NUMBER: US/09/172,952
;; CURRENT FILING DATE: 1998-10-14
;; NUMBER OF SEQ ID NOS: 33
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 33
;; LENGTH: 231
;; TYPE: PRT
;; ORGANISM: Yias-Hi
US-09-172-952-33

Query Match 9.8%; Score 128.5; DB 4; Length 231;
Best Local Similarity 27.8%; Pred. No. 7.1e-06;
Matches 50; Conservative 23; Mismatches 74; Indels 33; Gaps 9;
QY 35 YHLGWGTGGGISL--KHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKLK 92
DB 19 HHL--VTFWGNVSAIDREKXNVKPSGVYDVMTENDMVVDL-----FTGNIVEGNKK 72
QY 93 KSQCTPLFMNAY--TMRGAGAVIHTHSAKAVWATLLFPQGEFKITHQEMIKGIRKCT--- 147
DB 73 PSSDTPHLELYRQFPFHGGIVHTSHRATIWSQAGLDIIEVGTTHGDFYFTIPCTROM 132
QY 148 -----SGGYRYDDMLVVPPIENTPEKDKERMAHANNEYDPDSC-AVLVRHGVVYWG 201

Db 133 TTKKIGNY-----ELETGKVIVETFLSRGIE-----PDNIPAVLVHSHGPFANWK 178

RESULT 6

US-09-107-532A-4317

; Sequence 4317, Application US/09107532A

; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4317:

SEQUENCE CHARACTERISTICS:

LENGTH: 241 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...241

SEQUENCE DESCRIPTION: SEQ ID NO: 4317:

US-09-107-532A-4317

Query Match 9.5%; Score 125.5; DB 4; Length 241;

Best Local Similarity 26.1%; Pred. No. 1.6e-05;

Matches 59; Conservative 20; Mismatches 66; Indels 81; Gaps 11;

QY 45 GGISLKHGN-----EVIAPSGVQKERIQPEDMFVCDIN-----RQDISGPPASKKL 91

Db 30 GLAVKLWGVSEVDRELGVIVIKPSGVREYECMQADQMVVTDLSGNIIEED-----SL 81

QY 92 KKSQCTPLFMNAY--TWRGAGAVIHTHSAAYVATLLFPQOE---PKITHQEMIKGIRKCT 147

Db 82 KPSSDLPTHVLVYQTFEDITATHSTHSHVWMAQ--AGRDLPAYGTHRADFYGVKVFCT 139

QY 148 SGGYRYDDMLVPIIENTPPE-----KDLKERMAHAMNEYDPDSCAVLVRR 193

Db 140 R-----QLTKREYREAYEYVHTGNVIVETFKERKLDLP--NEVP---GVLVYG 180

QY 194 HGVVYWGTEWKA-----KTMCECYDILFD 218

Db 181 HGPFTWGDSPMKAVENSLILDBICLMAKENLINPNICEIPOVYILD 226

RESULT 7

US-09-107-532A-5562

; Sequence 5562, Application US/09107532A

; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5562:

SEQUENCE CHARACTERISTICS:

LENGTH: 233 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...233

SEQUENCE DESCRIPTION: SEQ ID NO: 5562:

US-09-107-532A-5562

Query Match 9.5%; Score 124.5; DB 4; Length 233;

Best Local Similarity 25.4%; Pred. No. 2e-05;

Matches 59; Conservative 27; Mismatches 101; Indels 45; Gaps 10;

QY 26 LIPELCKQFYHL-----GWVTGTGGGISLKHGNEIY--IAPSGVQKERIQPEDMFVCD 76

Db 2 MLEQLKEVYQANLDLPKHGLVKYTWGNVSADFDPETRYFVIKPSGVSYEEMTADDMVVVD 61

QY 77 INEQDISGPPASKKLKKSQCTPLFMNAY--TWRGAGAVIHTHSAAYVATLLFPQOEFKI 134

Db 62 LDNHIIEG-----KLNPSDDTPTHAVLYRSPQIGGVITHSTWATIIWAQGLDVPAMGT 116

QY 135 THQEMIKGIRKCT-----SGYRYDDMLVPIIENTPEEKDLKERMAHAMNEYPD 185

Db 117 THADTFYGVSPCARPLTQOEIDSG--YETGKV--IIEYFKKX-----ID 159

QY 186 SCA---VLVRHGVVYWGTEWKEKATMCCECYDYILFDIAVSMKMKGLDPTQLP 234

Db 160 PLATPGVLLHGHGPFTHGKDAQSAVMNAVVLDEVCKMNLFTQLNSFSEELP 211

RESULT 8
US-09-134-000C-6065
; Sequence 6065, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6065
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6065
Query Match 9.4%; Score 123.5; DB 4; Length 241;
Best Local Similarity 31.8%; Pred. No. 2.6e-05;
Matches 57; Conservative 17; Mismatches 82; Indels 23; Gaps 9;
QY 37 LGWVTGGG--LKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKKLKKS 94
DB 29 LGVVLKLTGNSVNSRLGIIIVKPSGVQKRIQPEDMFVCDINEQDISGPPASKKLKKS 84
QY 95 QCTPLFNNAY-TWRGAGAVHTHSKAAV---MATLLFPQGEFKITHQEMIKGIRKCTSGG 150
DB 85 SDFPLHLYLXQKPEIGALHTHSLNSVTWQAQGRALP--PYCTTHADAFYGAVPCTRA- 141
QY 151 YRYDMLVPIIENTPPE--KOLKERMAMANEYDSC-AVLVRHGVVYVWGTEWKA 206
DB 142 -----LSESEIKENVEETGKIVETF-HEQLDPLAVGLVYVGHGFTWGTPEKA 193
RESULT 9
US-08-472-534-2
; Sequence 2, Application US/08472534
; Patent No. 5919620
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard R
; APPLICANT: Martin, Denis
; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,534
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr, James F
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: Slovac-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000

TELEFAX: 212-596-9090
TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-534-2
Query Match 9.0%; Score 118; DB 2; Length 242;
Best Local Similarity 22.9%; Pred. No. 0.00011;
Matches 44; Conservative 31; Mismatches 83; Indels 34; Gaps 8;
QY 17 AQDKHPRFLPELCKQFYHLGWTGTGGISLKHGNEIYIA-PSGVQKRIQPEDMFVC 75
DB 2 SQDKLIREQICDVCHQWQLGVAANDGNVRLDEDTILATPTGISKSFITPEKLVKL 61
QY 76 DINEQDISP----PASKKLKKSQCTPLFMNAYTMR-GAGAVHTHSKAAV-MATLLRPG 129
DB 62 NLKGEILEABGDYCPSSSEIKMHIRC-----YEREDVRSVVHAHPPIATGFAHAHPL 114
QY 130 QBFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDLKERMAMANEY-PDSCA 188
DB 115 DYSILIESAIVGAIPITPPG-----VPSTWEVPE-----AITPYLPDHDV 155
QY 189 VLVRHGVVYVWG 200
DB 156 MLENHGALTVG 167
RESULT 10
US-08-926-842B-21
; Sequence 21, Application US/08926842B
; Patent No. 6030807
; GENERAL INFORMATION:
; APPLICANT: Sa-No. 6030807ueira, Isabel
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,842B
; FILING DATE: 10-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-089 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
US-08-926-842B-21

Query Match      8.6%; Score 113.5; DB 3; Length 240;
Best Local Similarity 26.4%; Pred. No. 0.00032;
Matches 47; Conservative 20; Mismatches 84; Indels 27; Gaps 7;

QY 40 VTGTGGGSL--KHGNEIYIAPSGVQKERIQEDMFVCDINEQDISGPPASKLKKSQCT 97
DB 22 VTLTGNVSAVDRERGVLIPSGVDYSIMTADDMMVVSL-----SGEVGHHKPPSDDT 77

QY 98 PLFNWY--TMRGAGAVIHTSK-AAVMATLLFPQGEFKITHQEMIKGIRKCTS-----148
DB 78 PTHRLLYQAFTIGGIVHTSRHATINAQAQPIATGCTTHADYFYGTIPCTRMTAEI 137

QY 149 GGYRYDDMLVPIENTPEEKDKERMAHAMNEYDPSCAVLRRHGVVVGSTWEKA 206
DB 138 NGEYEW---TGNVIVETFEKQGI-----DAAQMPGVLVSHGPPFANGKNAEDA 183

RESULT 11
US-09-489-039A-10152
; Sequence 10152, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10152
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10152

Query Match      8.3%; Score 109.5; DB 4; Length 225;
Best Local Similarity 22.9%; Pred. No. 0.00081;
Matches 50; Conservative 38; Mismatches 87; Indels 43; Gaps 10;

QY 27 IPELCKQFHLGWVTGTGGGSLKHGNEIYIAPSGVQKERIQEDMFVCDINEQDISGPP 86
DB 20 IITDCTLEMTRLGNOGTAGNSVRYQGMLTPTGIPYEKLKTEDKIVFIDADGQHEQGL 79

QY 87 ASKLLKKSQCTPLFNWYTW--GAGAVIHTSKAAVMATLL--PFGQBFKITHQEMIKG 142
DB 80 PSEWRFHQA-----AYQTRDAQVAVNHAVHCTAVSILNRPPIAHYMI-----125

QY 143 IRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYDPSCAVLRRHGVVVGST 202
DB 126 ---AAAGG---NSIFCAPYA--TFGTRELSHVAVALKHRK---ATLIQHGLIACEAS 173

QY 203 WKAKTMCSCYD-----YLFDTAVSMKKGLDPTQLPV 235
DB 174 LEKALWLAHEVEVLQAQLYLSLTAIT-----DP--VPV 203

RESULT 12
US-08-926-842B-20
; Sequence 20, Application US/08926842B
; Patent No. 6030807
; GENERAL INFORMATION:
; APPLICANT: Sa-No. 6030807ueira, Isabel
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,842B
; FILING DATE: 10-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-089 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
US-08-926-842B-20

Query Match      8.3%; Score 109.5; DB 3; Length 231;
Best Local Similarity 26.1%; Pred. No. 0.00084;
Matches 47; Conservative 19; Mismatches 83; Indels 31; Gaps 8;

QY 40 VTGTGGGSL--KHGNEIYIAPSGVQKERIQEDMFVCDINEQDISGPPASKLKKSQCT 97
DB 22 VTLTGNVSAVDRERGVLIPSGVDYSIMTADDMMVVSI---TGEVWEGAKKPPSDDT 77

QY 98 PLFNWY--TMRGAGAVIHTSKAAVMATLLFPQGEFK---ITHQEMIKGIRKCTS-----148
DB 78 PTHRLLYQAFTIGGIVHTSRHATINAQAQ--AGQSIPATGTTTHADYFYGTIPCTRMTDA 135

QY 149 --GGYRYDDMLVPIENTPEEKDKERMAHAMNEYDPSCAVLRRHGVVVGSTWEKA 206
DB 136 EINGEYEW---TGNVIVETFEKQGI-----DAAQMPGVLVSHGPPFANGKNAEDA 183

RESULT 13
US-08-840-683-11
; Sequence 11, Application US/08840683
; Patent No. 5821051
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-840-683-11

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGISLKHG-----NEIYIAPSGVQK 64
   |||||:|||||
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGFIYIAPSGCKR 49

RESULT 14
US-08-555-722-11
; Sequence 11, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-08-555-722-11

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGISLKHG-----NEIYIAPSGVQK 64
   |||||:|||||
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGFIYIAPSGCKR 49

RESULT 15
US-09-384-301-11
; Sequence 11, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-09-384-301-11

Query Match      8.2%; Score 108.5; DB 3; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGISLKHG-----NEIYIAPSGVQK 64
   |||||:|||||
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGFIYIAPSGCKR 49

Search completed: June 1, 2004, 13:35:58
Job time : 23 secs
```

```

Query Match      8.2%; Score 108.5; DB 2; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGISLKHG-----NEIYIAPSGVQK 64
   |||||:|||||
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGFIYIAPSGCKR 49

RESULT 15
US-09-384-301-11
; Sequence 11, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; US-09-384-301-11

Query Match      8.2%; Score 108.5; DB 3; Length 75;
Best Local Similarity 46.3%; Pred. No. 0.00023;
Matches 25; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

QY 16 GAQDKHPRFLPELCKQFYHLGWVTGGGISLKHG-----NEIYIAPSGVQK 64
   |||||:|||||
Db 2 GRQDKHPRYLPELANSF-----TIXAGSLGLEELAXSMGFIYIAPSGCKR 49

Search completed: June 1, 2004, 13:35:58
Job time : 23 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:33:26 ; Search time 48 Seconds
(without alignments)

1402.065 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQGCCSRPCGAQDK.....SMKKMGIDPTQLPVGNGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1317	100.0	241	12	US-10-381-710-2
2	1317	100.0	241	16	US-10-629-329A-4
3	1246.5	99.4	242	15	US-10-220-381-2
4	1239.5	94.1	242	16	US-10-629-329A-2
5	1235.5	93.8	242	12	US-10-381-710-4
6	550	41.8	256	12	US-10-424-599-157170
7	548.5	41.6	524	12	US-10-425-114-64486
8	540	41.0	497	12	US-10-425-114-65135
9	535	40.6	522	12	US-10-424-599-273717
10	535	40.6	540	12	US-10-425-114-46271
11	458.5	34.8	459	12	US-10-425-114-61505
12	405	30.8	533	12	US-10-425-114-57875
13	315	23.9	64	14	US-10-029-386-28983
14	222	16.9	59	14	US-10-106-698-5148
15	203	15.4	202	14	US-10-156-761-14187

16	201	15.3	212	12	US-10-282-122A-45375	Sequence 45375, A
17	195	14.8	204	12	US-10-282-122A-69836	Sequence 69836, A
18	172	13.1	204	12	US-10-282-122A-55860	Sequence 55860, A
19	164	12.5	205	12	US-10-282-122A-43563	Sequence 43563, A
20	161.5	12.3	241	14	US-10-156-761-13948	Sequence 13948, A
21	142	10.8	104	12	US-10-424-599-183312	Sequence 183312, A
22	139	10.6	231	10	US-09-557-796-32	Sequence 32, Appl
23	137	10.4	181	12	US-10-282-122A-46808	Sequence 46808, A
24	133	10.1	202	12	US-10-282-122A-59882	Sequence 59882, A
25	131	9.9	238	10	US-09-557-796-18	Sequence 18, Appl
26	128.5	9.8	231	10	US-09-557-796-33	Sequence 33, Appl
27	126.5	9.6	220	12	US-10-282-122A-49784	Sequence 49784, A
28	126	9.6	230	12	US-10-282-122A-77598	Sequence 77598, A
29	125	9.5	234	12	US-10-282-122A-74190	Sequence 74190, A
30	124.5	9.5	232	12	US-10-282-122A-57636	Sequence 57636, A
31	123.5	9.4	236	12	US-10-282-122A-57141	Sequence 57141, A
32	123	9.3	228	12	US-10-282-122A-75285	Sequence 75285, A
33	123	9.3	234	12	US-10-282-122A-74354	Sequence 74354, A
34	122.5	9.3	228	12	US-10-282-122A-53018	Sequence 53018, A
35	122.5	9.3	231	12	US-10-282-122A-77900	Sequence 77900, A
36	121	9.2	228	12	US-10-282-122A-76130	Sequence 76130, A
37	120	9.1	228	12	US-10-282-122A-56798	Sequence 56798, A
38	110.5	8.4	233	12	US-10-282-122A-51793	Sequence 51793, A
39	102.5	7.8	210	12	US-10-282-122A-58385	Sequence 58385, A
40	99.5	7.6	212	12	US-10-282-122A-48161	Sequence 48161, A
41	99	7.5	213	12	US-10-282-122A-50759	Sequence 50759, A
42	98.5	7.5	217	12	US-10-282-122A-61813	Sequence 61813, A
43	98	7.4	228	12	US-10-282-122A-73117	Sequence 73117, A
44	98	7.4	307	12	US-10-335-977-9353	Sequence 9353, Ap
45	98	7.4	315	12	US-10-335-977-9354	Sequence 9354, Ap

ALIGNMENTS

RESULT 1
US-10-381-710-2
; Sequence 2, Application US/10381710
; Publication No. US20040052789A1
; GENERAL INFORMATION:
; APPLICANT: SHA, Shiken et al.
; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME
; FILE REFERENCE: 0230-0198P
; CURRENT APPLICATION NUMBER: US/10/381,710
; CURRENT FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Mouse macrophage cell RAW 264.7
US-10-381-710-2

Query Match 100.0%; Score 1317; DB 12; Length 241;
Best Local Similarity 100.0%; Pred. No. 1e-133;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSGCQAQGCCSRPCGAQDKHEPRFLPELCKQFYHLGWTGTGGISLKHGNEIYIAPS	60
Db	1	MSGCQAQGCCSRPCGAQDKHEPRFLPELCKQFYHLGWTGTGGISLKHGNEIYIAPS	60
Qy	61	GVQKERIQEDMFVCDINEQDISGPPASKLKKSQCTPLFNAYTMRGAGAVIHTSKAA	120
Db	61	GVQKERIQEDMFVCDINEQDISGPPASKLKKSQCTPLFNAYTMRGAGAVIHTSKAA	120
Qy	121	VMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIENTPREKDKERMAHAM	180
Db	121	VMATLLFPQGFKITHQEMIKIRKCTSGGYRYDDMLVPIENTPREKDKERMAHAM	180
Qy	181	NEYPDSCAVLRHGGVYWGVTWEKATMCFCYDLFDIAVSMKKMGIDPTQLPVGNGI	240
Db	181	NEYPDSCAVLRHGGVYWGVTWEKATMCFCYDLFDIAVSMKKMGIDPTQLPVGNGI	240

```
QY 241 V 241
Db 241 V 241

RESULT 2
US-10-629-329A-4
; Sequence 4, Application US/10629329A
; Publication No. US20040086848A1
; GENERAL INFORMATION:
; APPLICANT: DARNAY, BRYANT G.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING POLYNUCLEOTIDES AND
; TITLE OF INVENTION: POLYPEPTIDES OF RANK-ASSOCIATED INHIBITOR (RAIN)
; FILE REFERENCE: UTSC:761US
; CURRENT APPLICATION NUMBER: US/10/629,329A
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: 60/399,205
; PRIOR FILING DATE: 2002-07-29
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-629-329A-4

Query Match 100.0%; Score 1317; DB 16; Length 241;
Best Local Similarity 100.0%; Pred. No. 1e-133; Indels 0; Gaps 0;
Matches 241; Conservative 0; Mismatches 0;

QY 1 MSGCQAQGCCSRPCGAQDKHEHPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAPS 60
Db 1 MSGCQAQGCCSRPCGAQDKHEHPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAPS 60
QY 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTHSKAA 120
Db 61 GVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTHSKAA 120
QY 121 VMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180
Db 121 VMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHAM 180
QY 181 NEYPDSCAVLVRHGVVWGTWEKATMCCECYDYLFDIAVSMKMGDLPTQLPVGENGI 240
Db 181 NEYPDSCAVLVRHGVVWGTWEKATMCCECYDYLFDIAVSMKMGDLPTQLPVGENGI 240
QY 241 V 241
Db 241 V 241

RESULT 3
US-10-220-381-2
; Sequence 2, Application US/10220381
; Publication No. US20030207430A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Dying Aina M.
; APPLICANT: BANDMAN, Olga
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LAL, Preeti
; APPLICANT: BURFORD, Neil
; APPLICANT: BAUGHN, Mariah R.
; TITLE OF INVENTION: HUMAN ENZYME MOLECULES
; FILE REFERENCE: PF-0763 PCT
; CURRENT APPLICATION NUMBER: US/10/220,381
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 242

QY 241 V 241
Db 241 V 241

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030207430A1 2116390CD1
US-10-220-381-2

Query Match 94.1%; Score 1239,5; DB 16; Length 242;
Best Local Similarity 93.8%; Pred. No. 2.5e-125;
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCQA-QGCCSRPCGAQDKHEHPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAP 59
Db 1 MSGCDAGEGDCSRPCGAQDKHEHPRFLIPELCKQFYHLGWVTGTGGGISLKHGDEIYIAP 60
QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTHSKA 119
Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTHSKA 120
QY 120 AVMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
Db 121 AVMATLLPFGREFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKORMAHA 180
QY 180 MNEYPDSCAVLVRHGVVWGTWEKATMCCECYDYLFDIAVSMKMGDLPTQLPVGENG 239
Db 181 MNEYPDSCAVLVRHGVVWGTWEKATMCCECYDYLFDIAVSMKMGDLPSQLPVGENG 240

RESULT 4
US-10-629-329A-2
; Sequence 2, Application US/10629329A
; Publication No. US20040086848A1
; GENERAL INFORMATION:
; APPLICANT: DARNAY, BRYANT G.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING POLYNUCLEOTIDES AND
; TITLE OF INVENTION: POLYPEPTIDES OF RANK-ASSOCIATED INHIBITOR (RAIN)
; FILE REFERENCE: UTSC:761US
; CURRENT APPLICATION NUMBER: US/10/629,329A
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: 60/399,205
; PRIOR FILING DATE: 2002-07-29
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-629-329A-2

Query Match 94.1%; Score 1239,5; DB 16; Length 242;
Best Local Similarity 93.8%; Pred. No. 2.5e-125;
Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCQA-QGCCSRPCGAQDKHEHPRFLIPELCKQFYHLGWVTGTGGGISLKHGNEIYIAP 59
Db 1 MSGCDAGEGDCSRPCGAQDKHEHPRFLIPELCKQFYHLGWVTGTGGGISLKHGDEIYIAP 60
QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTHSKA 119
Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFNNAYTMRGAGAVIHTHSKA 120
QY 120 AVMATLLPFGQEFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
Db 121 AVMATLLPFGREFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKORMAHA 180
QY 180 MNEYPDSCAVLVRHGVVWGTWEKATMCCECYDYLFDIAVSMKMGDLPTQLPVGENG 239
Db 181 MNEYPDSCAVLVRHGVVWGTWEKATMCCECYDYLFDIAVSMKMGDLPSQLPVGENG 240
```

Matches 121; Conservative 33; Mismatches 74; Indels 26; Gaps 9;

QY 1 MSGCOAQGCCSRPC---GAQDKHPRFLIPELCKQFVHLGWVTGGGSIK-HGNE-- 54

Db 1 MAATATADSVSSQAYLEGNVKE--TRILISELCRHFYSLGWVSGSITIKVHDDIS 59

QY 55 -----IYIAPSGVOKERIQPEDMFVCDINEQDISGPPAS-----KKLKSOCTPLFMNAYT 105

Db 60 KPHQILMSPSGVOKERMEPEMDYVLSHTGVSLSAPSPKPYPHKPKSCDCGFLPMKAYE 119

QY 106 MRGAGAVIHTSKAAVMATLLFP--QOEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIE 164

Db 120 MCDAGAVIHTSGIESCLVTLNPLAKEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIE 173

QY 165 NTPPEKOLKERMAMHANEYPDSCAVLVRHGHVYVWGETWEKAKTWCBCYDYLFDIAVSMK 224

Db 174 NTAYEVELTESLAKAIEAYPKTTAVLVRNHGIYVWGSNISAKTQAECHYHLYLDAIKUH 233

QY 225 KMGLD---PTQLPV 235

Db 234 QLGLDWSTPNHVPI 247

RESULT 7

US-10-425-114-64486

; Sequence 64486, Application US/10425114

; Publication No. US2004003488A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 64486

; LENGTH: 524

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3060-104-F8_FLI.pep

US-10-425-114-64486

Query Match 41.6%; Score 548.5; DB 12; Length 524;

Best Local Similarity 45.6%; Pred. No. 4.1e-50;

Matches 115; Conservative 36; Mismatches 68; Indels 33; Gaps 6;

QY 2 SGCOAQGCCSRPCGAQDKE-----HPRFLIPELCKQFVHLGWVTGGGSIKXHG 52

Db 11 SGCS-----CEAAVGAMASEAYLEGAPVREARELVABELCRHFYAGWVVTGGGSIYKVN 65

QY 53 NE-----IYIAPSGVOKERIQPEDMFVCDINEQDISGPPA-----SKLKSOCTPLF 100

Db 66 DPTVPLADRLIVMSPSGVOKERVAEDMYMAADGKVLISAPVAKPWPKNPKTDCAPLF 125

QY 101 MNATYMGAGAVIHTSKAAVMATLLFP--QOEFKITHQEMIKGIRKCTSGGYRYDDMLV 159

Db 126 MKAYLMRGAGAVIHTSGIETCTIATMLIPGAKEFVTHQEMIKGIRKCTSGGYRYDDMLV 179

QY 160 VPIIENTPEEKOLKERMAMHANEYPDSCAVLVRHGHVYVWGETWEKAKTWCBCYDYLFDI 219

Db 180 IPIIENTPEYVELTDSLSEALAAYPKATAVLVRNHGIYVWGSNISAKTQAECHYHLYLDA 239

QY 220 AVSMKMGGLDPT 231

Db 240 CIKLYQLGIDWT 251

Matches 240 IV 241

Db 241 IV 242

RESULT 5

US-10-381-710-4

; Sequence 4, Application US/10381710

; Publication No. US20040052789A1

; GENERAL INFORMATION:

; APPLICANT: SHA, Shiken et al.

; TITLE OF INVENTION: NOVEL PROTEINS, GENES ENCODING THEM AND METHOD OF USING THE SAME

; FILE REFERENCE: 0230-0198P

; CURRENT APPLICATION NUMBER: US/10/381,710

; CURRENT FILING DATE: 2003-09-16

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 4

; LENGTH: 242

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-381-710-4

Query Match 93.8%; Score 1235.5; DB 12; Length 242;

Best Local Similarity 93.8%; Pred. No. 6.8e-125;

Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCOA-QGCCSRPCGAQDKEHPRFLIPELCKQFVHLGWVTGGGSIKXGNEIYIAP 59

Db 1 MSGCDAMEGDCSRRCGAQDKEHPRFLIPELCKQFVHLGWVTGGGSIKXGDEIYIAP 60

QY 60 SGVOKERIQPEDMFVCDINEQDISGPPASKKLKSQCTPLFMNAYTMRGAGAVIHTSKA 119

Db 61 SGVOKERIQPEDMFVCDINEQDISGPPASKKLKSQCTPLFMNAYTMRGAGAVIHTSKA 120

QY 120 AVMATLLFPQOEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEEKOLKERMAMA 179

Db 121 AVMATLLFPQOEFKITHQEMIKGIRKCTSGGYRYDDMLVVPPIENTPEEKOLKERMAMA 180

QY 180 MNEYPDSCAVLVRHGHVYVWGETWEKAKTWCBCYDYLFDIAVSMKMGGLDPTQLPVGNG 239

Db 181 MNEYPDSCAVLVRHGHVYVWGETWEKAKTWCBCYDYLFDIAVSMKMGGLDPTQLPVGNG 240

QY 240 IV 241

Db 241 IV 242

RESULT 6

US-10-424-599-157170

; Sequence 157170, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 157170

; LENGTH: 256

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_112946C.1.pep

US-10-424-599-157170

Query Match 41.8%; Score 550; DB 12; Length 256;

Best Local Similarity 47.6%; Pred. No. 1e-50;


```
RESULT 8
US-10-425-114-65135
; Sequence 65135, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65135
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17103D03_FLI.pep
US-10-425-114-65135

Query Match 41.0%; Score 540; DB 12; Length 497;
Best Local Similarity 48.5%; Pred. No. 3.2e-49;
Matches 111; Conservative 37; Mismatches 61; Indels 20; Gaps 5;

QY 16 GAQDKEHRELPELCKQFVHLGWVTGGGSIK-HGNE-----IYIAPSGVQKRIQPEDMFVC 67
DB 3 GAVPRE-ARELVLCRHFYQAQWVTGGTGGTIVKYNDAVPLADRLIYVSPSGVQKERM 61
QY 68 QPEDMFVCDINEQDISGPPA-----SKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMA 123
DB 62 VAEDMYVMAADGKVSAPVAKPWPKPCKTDCAPLFMAKAYLMRGAGAVIHTSHGIETCIA 121
QY 124 TLLPFG-QEFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNE 182
DB 122 TMLIPGAKEFRVTHMEMIKIGK-----HGYHDELVIPIENTPYEYELTDSLSEIAA 175
QY 183 YPDSCAVLVRHGVYVWGETWEKAKTMCCEYDYLFDIAVSMKKMGDPT 231
DB 176 YPRATAVLVRNHGIVYWGSEWINAKTOACBCHYLLDACIKLYQLGIDWT 224

RESULT 9
US-10-424-599-273717
; Sequence 273717, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 273717
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_89188C.1.pep
US-10-424-599-273717

Query Match 40.6%; Score 535; DB 12; Length 522;
Best Local Similarity 50.4%; Pred. No. 1.2e-48;
Matches 115; Conservative 28; Mismatches 63; Indels 22; Gaps 7;

US-10-425-114-46271
; Sequence 46271, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46271
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701207771_FLI.pep
US-10-425-114-46271

Query Match 40.6%; Score 535; DB 12; Length 540;
Best Local Similarity 50.4%; Pred. No. 1.2e-48;
Matches 115; Conservative 28; Mismatches 63; Indels 22; Gaps 7;

QY 24 RFLIPELCKQFVHLGWVTGGGSIK-HGNE-----IYIAPSGVQKRIQPEDMFVC 75
DB 47 RALMAELCRHFYTLGWVTGGTGGTSGISMKVHDDSTPRPQQLILMAPSGVQKERMEDMYVL 106
QY 76 DINEQDISGPPAS-----KKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMAI LFP-GQ 130
DB 107 SHSGSVLSAPSPPKWPHPKPCSDCDPLFKKAYEMRDAVAFHSHGIESCLVTMINPLSK 166
QY 131 EFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVL 190
DB 167 EFRITHMEMIKGIK---GHGY---DELVPIIENTAYEYQLTESFAKAIEDYPKATAVL 220
QY 191 VRHGVYVWGETWEKAKTMCCEYDYLFDIAVSMKKMGD---PTQLPV 235
DB 221 VRNHGVFVWGDWSISAKTQSECHYLFDAALKLHOMGLDWSPTNHPGI 268

RESULT 11
US-10-425-114-61505
; Sequence 61505, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

```
QY 24 RFLIPELCKQFVHLGWVTGGGSIK-HGNE-----IYIAPSGVQKRIQPEDMFVC 75
DB 29 RALMAELCRHFYTLGWVTGGTGGTSGISMKVHDDSTPRPQQLILMAPSGVQKERMEDMYVL 88
QY 76 DINEQDISGPPAS-----KKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMAI LFP-GQ 130
DB 89 SHSGSVLSAPSPPKWPHPKPCSDCDPLFKKAYEMRDAVAFHSHGIESCLVTMINPLSK 148
QY 131 EFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVL 190
DB 149 EFRITHMEMIKGIK---GHGY---DELVPIIENTAYEYQLTESFAKAIEDYPKATAVL 202
QY 191 VRHGVYVWGETWEKAKTMCCEYDYLFDIAVSMKKMGD---PTQLPV 235
DB 203 VRNHGVFVWGDWSISAKTQSECHYLFDAALKLHOMGLDWSPTNHPGI 250
```

```
RESULT 10
US-10-425-114-46271
; Sequence 46271, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46271
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701207771_FLI.pep
US-10-425-114-46271

Query Match 40.6%; Score 535; DB 12; Length 540;
Best Local Similarity 50.4%; Pred. No. 1.2e-48;
Matches 115; Conservative 28; Mismatches 63; Indels 22; Gaps 7;

QY 24 RFLIPELCKQFVHLGWVTGGGSIK-HGNE-----IYIAPSGVQKRIQPEDMFVC 75
DB 47 RALMAELCRHFYTLGWVTGGTGGTSGISMKVHDDSTPRPQQLILMAPSGVQKERMEDMYVL 106
QY 76 DINEQDISGPPAS-----KKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVMAI LFP-GQ 130
DB 107 SHSGSVLSAPSPPKWPHPKPCSDCDPLFKKAYEMRDAVAFHSHGIESCLVTMINPLSK 166
QY 131 EFKITHQEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVL 190
DB 167 EFRITHMEMIKGIK---GHGY---DELVPIIENTAYEYQLTESFAKAIEDYPKATAVL 220
QY 191 VRHGVYVWGETWEKAKTMCCEYDYLFDIAVSMKKMGD---PTQLPV 235
DB 221 VRNHGVFVWGDWSISAKTQSECHYLFDAALKLHOMGLDWSPTNHPGI 268
```

```
RESULT 11
US-10-425-114-61505
; Sequence 61505, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53313)B
 CURRENT APPLICATION NUMBER: US/10/425,114
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 73128
 SEQ ID NO 61505
 LENGTH: 459
 TYPE: PRT
 ORGANISM: Zea mays
 FEATURE:
 OTHER INFORMATION: Clone ID: LIB143-005-F3_FLI.pap
 US-10-425-114-61505

Query Match 34.8%; Score 458.5; DB 12; Length 459;
 Best Local Similarity 50.0%; Pred. No. 1.8e-40;
 Matches 91; Conservative 31; Mismatches 49; Indels 11; Gaps 3;

QY 55 IVIAPSGVQKRIQPEDMFVCDINEQDISGPPA----SKLKKSQCTPLFMAYTMRGAG 110
 DB 11 IVMSPSGVQKRWAEVDYVMAADGKVLGAPVAKPWNPKKCTDCAPLFMKAYLMRGAG 70
 QY 111 AVITHSKAAVMATLFPG-QEPKITHQEMIKGIRKCTSGGYRYDDMLVPIIENPTPEE 169
 DB 71 AVITHSHGIECTIATMLIPGAKERFVTHMEMIKIGK-----HGVDHDELVIPIIENPTPE 124
 QY 170 KOLKERMAMANEYDSCAVLVRHGHVYVWGETWEKAKTWCBCYDYLFDIAVSMKMGGLD 229
 DB 125 YELTDSLSEAIAYPRKATVLRVNHGIVWGESWINAKTQACVHYLLDACIKLYQLGID 184
 QY 230 PT 231
 DB 185 WT 186

RESULT 12
 US-10-425-114-57875
 Sequence 57875, Application US/10425114
 Publication No. US20040034888A1
 GENERAL INFORMATION:
 APPLICANT: Liu, Jingdong
 APPLICANT: Zhou, Yihua
 APPLICANT: Kovacic, David K.
 APPLICANT: Screen, Steven E.
 APPLICANT: Tabaska, Jack E.
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53313)B
 CURRENT APPLICATION NUMBER: US/10/425,114
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 73128
 SEQ ID NO 57875
 LENGTH: 533
 TYPE: PRT
 ORGANISM: Zea mays
 FEATURE:
 OTHER INFORMATION: Clone ID: UC-ZWFLM017051F08_FLI.pap
 US-10-425-114-57875

Query Match 30.8%; Score 405; DB 12; Length 533;
 Best Local Similarity 40.0%; Pred. No. 1.4e-34;
 Matches 88; Conservative 32; Mismatches 64; Indels 36; Gaps 6;

QY 42 GTGGG-----ISLKHGNEIYIAPSGVQKRI--QPEDMF-----VCD- 76
 DB 47 GAGGAGSALLPAGMGHGRHROHFGORPRRARRPHRYVTLRCAGEDSGSRHVCVG 106
 QY 77 ----INEQDISGPPASKLKKSQCTPLFMAYTMRGAGAVITHSKAAVMATLFPG-QE 131
 DB 107 CRWEVLSAPVAKPWNPKKCTDCAPLFMKAYLMRGAGAVITHSHGIECTIATMLIPGAK 166

QY 132 FKITHQEMIKGIRKCTSGGYRYDDMLVPIIENPTPEKDKERMAMANEYDSCAVLV 191
 DB 167 FRVTHMEMIKIGK-----HGVDHDELVIPIIENPTPEYELTDSLSEAIAYPRKATV 220
 QY 192 RPHGVYVWGETWEKAKTWCBCYDYLFDIAVSMKMGGLDPT 231
 DB 221 RNHGIVWGESWINAKTQACVHYLLDACIKLYQLGIDWT 260

RESULT 13
 US-10-029-386-28983
 Sequence 28983, Application US/10029386
 Publication No. US20030194704A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
 TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 FILE REFERENCE: AEOMICA-X-2
 CURRENT APPLICATION NUMBER: US/10/029,386
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 34288
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 28983
 LENGTH: 64
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO CHR11.1
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.86
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.86
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.67
 OTHER INFORMATION: SWISSPROT HIT: P47095, EVALUATION 1.00e-11
 US-10-029-386-28983

Query Match 23.9%; Score 315; DB 14; Length 64;
 Best Local Similarity 98.3%; Pred. No. 3.6e-26;
 Matches 57; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 152 YRYDDMLVPIIENPTPEKDKERMAMANEYDSCAVLVRHGHVYVWGETWEKAKTM 209
 DB 7 YRYDDMLVPIIENPTPEKDKERMAMANEYDSCAVLVRHGHVYVWGETWEKAKTM 64

RESULT 14
 US-10-106-698-5148
 Sequence 5148, Application US/10106698
 Publication No. US20030109690A1
 GENERAL INFORMATION:
 APPLICANT: Ruben et al.
 TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
 FILE REFERENCE: PA005F1
 CURRENT APPLICATION NUMBER: US/10/106,698
 CURRENT FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: PCT/US00/26524
 PRIOR FILING DATE: 2000-09-28
 PRIOR APPLICATION NUMBER: US 60/157,137
 PRIOR FILING DATE: 1999-09-29
 PRIOR APPLICATION NUMBER: US 60/163,280
 PRIOR FILING DATE: 1999-11-03
 NUMBER OF SEQ ID NOS: 8564
 SOFTWARE: PatentIn Ver. 3.0
 SEQ ID NO 5148
 LENGTH: 59
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-106-698-5148

Query Match 16.9%; Score 222; DB 14; Length 59;
 Best Local Similarity 95.2%; Pred. No. 3.6e-16;
 Matches 40; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 200 GETWEKAKTMCCEYDYLFDIAVSMKMGGLDPTQLPVGENGIV 241
Db 18 GETWEKAKTMCCEYDYLFDIAVSMKMGGLDPTQLPVGENGIV 59

RESULT 15
US-10-156-761-14187
; Sequence 14187, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14187
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14187

Query Match 15.4%; Score 203; DB 14; Length 202;
Best Local Similarity 27.7%; Pred.No. 2.4e-13;
Matches 54; Conservative 40; Mismatches 83; Indels 18; Gaps 5;
QY 33 QFVHLGWVTGGGSLKHGNE---IYIAPSGYQKRIQPEDMFVCDINEQDISGPPASK 89
Db 23 RFASFGMRGTSGNLSVLSRDLRLAVTASGHDKGELTADVVLVDGGA VAGGRPSA 82
QY 90 KLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVNMATLLFGQEFKITHQEMIKGIRKCTSG 149
Db 83 E-----AALHARVAALTGAGAVVHVTVASVANGREPG-GIAFRDLEMLKGV----- 129
QY 150 GYRYDDMLVVPITIENTPEEKDKERMAHAMNEYDSCAVLVRHGVYVWGETWEKAKTM 209
Db 130 GLPAHDVEVTLPIVANSQDMKVLGDRLEARD--PRMPAVVAVAGHGLYVWGENPROARHH 187
QY 210 CECYDYLFDIAVSMK 224
Db 188 TEVWEMLELELTQR 202

Search completed: June 1, 2004, 13:36:59
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:20:11 ; Search time 17 Seconds
(without alignments)

738.171 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSCCQAGGCCSPCGAQDK.....SMKKMGLDPTQLPVGENGIV 241

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	507	38.5	244	1	YJZ4 YEAST
2	188.5	14.3	208	1	P67788 aquifex aeo
3	152	11.5	181	1	YJ18 METUA
4	139	10.6	231	1	SGBE_ECOLI
5	132	10.0	242	1	SGRE_MYCPN
6	128.5	9.8	231	1	SGBE_HAEIN
7	128	9.7	229	1	ARAD_BACSU
8	125.5	9.5	228	1	ARAD_BACST
9	124	9.4	231	1	ARAD_BACDH
10	120	9.1	228	1	SGAE_ECOLI
11	113.5	8.6	231	1	ARAD_SALTY
12	110	8.4	216	1	FUCA_HAEIN
13	109.5	8.3	231	1	ARAD_ECOLI
14	108.5	8.2	215	1	FUCA_ECOLI
15	102.5	7.8	210	1	YGBL_HAEIN
16	98	7.4	284	1	PNK_HELPU
17	97	7.4	706	1	ADGG_HUMAN
18	96	7.3	705	1	ADGG_RAT
19	93	7.1	706	1	ADGG_MOUSE
20	92.5	7.0	2194	1	GLSN_MEDSA
21	91.5	6.9	212	1	YGBL_ECOLI
22	89.5	6.8	304	1	RGGC_VIBCH
23	89	6.8	536	1	FYN_XIPHE
24	88.5	6.7	284	1	PNK_HELPU
25	86.5	6.6	726	1	ADDB_HUMAN
26	86	6.5	352	1	Z185_MOUSE
27	85	6.5	304	1	RDGC_VIBPA
28	85	6.5	859	1	NIA_FICAN
29	83.5	6.3	725	1	ADDB_MOUSE
30	83.5	6.3	725	1	ADDB_RAT
31	82.5	6.3	1729	1	DME_ARATH
32	80.5	6.1	304	1	RDGC_VIBVU
33	80	6.1	535	1	YRK_CHICK

34	80	6.1	536	1	FYN_XENLA
35	79.5	6.0	196	1	UHPA_SALTY
36	79.5	6.0	501	1	DGRE_DROME
37	78.5	6.0	231	1	Y493_RICPR
38	78.5	6.0	689	1	YUXG_BACSU
39	78	5.9	382	1	VA23_VARV
40	78	5.9	1317	1	N145_YEAST
41	77.5	5.9	271	1	CPHB_SYNY3
42	77.5	5.9	1650	1	VIT6_CAEEL
43	77	5.8	436	1	RHO_AQUAE
44	77	5.8	643	1	RO60_CAEEL
45	77	5.8	1043	1	PLID_MOUSE

ALIGNMENTS

RESULT 1

ID	YJZ4 YEAST	STANDARD;	PRT;	244 AA.
AC	P47035;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Hypothetical 27.4 kDa protein in MER2-BNAL intergenic region.			
GN	YJRO24C OR J1545 OR YJUR83.18.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96109930; PubMed=8619316;			
RA	Zagulski M., Babinska B., Gromadka R., Migdalski A., Rytka J.,			
RA	Sulicka J., Herbert C.J.;			
RT	"The sequence of 24.3 kb from chromosome X reveals five complete open			
RT	reading frames, all of which correspond to new genes, and a tandem			
RT	insertion of a Ty1 transposon.";			
RL	Yeast 11:1179-1186(1995).			
RN	[2]			
RP	SEQUENCE OF 44-244 FROM N.A.			
RC	STRAIN=S288c / FY1679;			
RA	de Haan M., Smits P.H.M., Grivell L.A.;			
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; Z49524; CAA89549.1; -			
DR	EMBL; X87297; CAA60719.1; -			
DR	EMBL; X87611; CAA60947.1; -			
DR	PIR; S57042; S57042.			
DR	GermOnline; 141861; -			
DR	SGD; S0003785; YJRO24C			
DR	InterPro; IPR001303; Aldolase_II_N.			
DR	Pfam; PF00596; Aldolase_II; 1.			
KW	Hypothetical protein.			
SK	SEQUENCE 244 AA; 27427 MW; C41FAD142D677A00 CRC64;			

Query Match 38.5%; Score 507; DB 1; Length 244;

Best Local Similarity 46.0%; Pred. No. 1.2e-37;

Matches 108; Conservative 36; Mismatches 67; Indels 24; Gaps 7;

QY 14 PCGADQKHEHREFLPELCKQFVHLGWVTGTGGISLK--HGNEIYIAPSGVQKRIQPED 71

DB 13 PC-----HPANLICTLCKQFFHNWCTGTGGISIKDPNTNYIYLAISGVQKRIQPED 66

QY 72 MFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTMRGAGAVIHTHSKAAVNTLLFPQGE 131

```

Db 67 LFVMDAQTEVLRSP--KLYPSACTPLFLACYQKNAGAIHHSQNAVICSLF-GDE 123
QY 132 FKITHQEMIKGI-----RKTSGGYRYDDMLVVPPIIENPTEPKDKLKERMAHANE 182
Db 124 FRIANIEQIKAIPSCKVDPVTKKPWALSFF---DTLKPIIENMAHEDELIDDLHKTFKD 180
QY 183 YPDSCAVLVRRHGVVWGETWEKAKTMCCEVDYLFDAVSMKMGCLDPTQLPVGE 237
Db 181 YPDTCAVIVRRHGIFVWGPFTDKAKIFNEAIDYLMELAIKTQMGF-PPDCGIGE 234

RESULT 2
YU79_AQUAE STANDARD; PRT; 208 AA.
AC O67788;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical aldolase class II protein AQ_1979.
GN AQ_1979.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -!- SIMILARITY: Belongs to the aldolase class II family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000766; AAC07751.1; -
CC PIR; H70469; H70469.
CC InterPro; IPR001303; Aldolase II_N.
CC DR InterPro; IPR004782; Fuca.
CC DR Pfam; PF00596; Aldolase II; 1.
CC DR TIGRFAMS; TIGR01086; fuca; 1.
CC KW Hypothetical protein; Zinc; Complete proteome.
CC FT METAL 80 80 ZINC (BY SIMILARITY).
CC FT METAL 99 99 ZINC (BY SIMILARITY).
CC FT METAL 101 101 ZINC (BY SIMILARITY).
CC FT METAL 174 174 ZINC (BY SIMILARITY).
CC SQ SEQUENCE 208 AA; 329550 MW; 3295525CED17344 CRC64;

Query Match 14.3%; Score 188.5; DB 1; Length 208;
Best Local Similarity 24.6%; Pred. No. 1.4e-09;
Matches 52; Conservative 39; Mismatches 93; Indels 27; Gaps 5;

QY 29 ELCKQF-----YHLGWVTGGGSLKHGNE-IYIAPSGVOKERIQPDMEF 73
Db 4 ELFKFKSEKVEEIIIEAGRIILHSRGWVPATSGNISAKVSEEVIAITAGKHKGLKLTPEDIL 63
QY 74 VCDINEQDI-SGPPASKKLKKSQCTPLFMNATYMRGAGAVIHTHSKAAMVATLLFPQGEF 132
Db 64 LIDYEGRPVGGKGSAAETLLHTTYKLFPE-----VNAVVHTSPNATVISIVEKKDFV 117
QY 133 KITHQEMIKGIRKCTSGGYRYDDMLVVPPIIENPTEPKDKLKERMAHANEYPDSCAVILVR 192
Db 118 ELEDYELLKAF-----PDITHVKIKIPFPNEQNIPLLAKEVENYFKTSEDYKGLIR 172

```

```

QY 193 RHGVVWGETWEKAKTMCCEVDYLFDAVSM 223
Db 173 GHGLYTWGRSMEALIHTEALEFIFECLEKL 203

RESULT 3
YE18_METJA STANDARD; PRT; 181 AA.
AC Q58813;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical aldolase class II protein MJI418.
GN MJI418.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.D., Sadow P.W., Hanna M.C.,
RA Utterback T.R., Kelley J.M., Peterson J.F., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -!- SIMILARITY: Belongs to the aldolase class II family. Arab/fuca
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67582; AAB99428.1; -
CC PIR; A64477; A64477.
CC HSSP; P11550; LDZX.
CC TIGR; MJI418; -.
CC InterPro; IPR001303; Aldolase II_N.
CC DR Pfam; PF00596; Aldolase II; 1.
CC KW Hypothetical protein; Zinc; Complete proteome.
CC FT METAL 68 68 ZINC (BY SIMILARITY).
CC FT METAL 87 87 ZINC (BY SIMILARITY).
CC FT METAL 89 89 ZINC (BY SIMILARITY).
CC FT METAL 147 147 ZINC (BY SIMILARITY).
CC SQ SEQUENCE 181 AA; 20470 MW; E5F3BF13722145B0 CRC64;

Query Match 11.5%; Score 152; DB 1; Length 181;
Best Local Similarity 24.1%; Pred. No. 2e-06;
Matches 42; Conservative 38; Mismatches 66; Indels 28; Gaps 5;

QY 29 ELCKQFHYLGVWVTGGGSLKHGNEIYIAPSGVOKERIQPDMEFVCDINEQDISPPAS 88
Db 8 KICRKLKYVGVVSGNVSVKGDGKXILYTPGSLGLFLKEDDIAEMDLGDNVIRKGPST 67
QY 89 KLUKKSQCTPLFMNATYMRG-AGAVIHTHSKAAMVATLLFPQGEFKITHQEMIKGRKCT 147
Db 68 EK-----NLHLMYRKNDINAIHTHSLISTFLSTI--NKEIELLTEGKIFLKKI- 117
QY 148 SGGYRYDDMLVVPPIIENPTEPKDKLKERMAHANEYPDSCAVILVRRHGVVWGE 201

```

Db 118 --GYVDYEEAGSLKAEETAKR-----DEDVILKNHGVVCLGK 154

RESULT 4

```
SGBE_ECOLI      STANDARD;      PRT;      231 AA.
ID  SGBE_ECOLI
AC  P37680;
DT  01-OCT-1994 (Rel. 30, Created)
DT  01-OCT-1994 (Rel. 30, Last sequence update)
DE  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Probable sugar isomerase sgbe (EC 5.1.-.-).
GN  SGBE OR B3583.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Escherichia.
OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RC  MEDLINE=94316500; PubMed=8041620;
RA  Sofia H.J., Burland V., Daniels D.L., plunkett G. III, Blattner F.R.;
RT  "Analysis of the Escherichia coli genome. V. DNA sequence of the
RL  region from 76.0 to 81.5 minutes.";
RL  Nucleic Acids Res. 22:2576-2586(1994).
RN  [2]
RP  DISCUSSION OF SEQUENCE.
RA  Raizer J., Charbit A., Reizer A., Saier M.H. Jr.;
RT  "Novel phosphotransferases system genes revealed by bacterial genome
RT  analysis: operons encoding homologues of sugar-specific permease
RT  domains of the phosphotransferase system and pentose catabolic
RT  enzymes.";
RL  Genome Sci. Technol. 1:53-75(1996).
CC  -1- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE.
CC  -1- COPACTOR: Binds 1 zinc ion per molecule (Potential).
CC  -1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC  subfamily.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; U00039; AAB18560.1; -.
CC  EMBL; AE000435; AAC76607.1; -.
CC  PIR; S47804; S47804.
CC  EcoGene; EG12287; sgbe.
CC  InterPro; IPR001303; Aldolase II N.
CC  Pfam; PF00596; Aldolase II; 1.
CC  Isomerase; Zinc; Complete proteome.
CC  METAL 76 76 ZINC (BY SIMILARITY).
CC  METAL 95 95 ZINC (BY SIMILARITY).
CC  METAL 97 97 ZINC (BY SIMILARITY).
CC  METAL 171 171 ZINC (BY SIMILARITY).
CC  SEQUENCE 231 AA; 25561 MW; FAFFAD7EC2A80B3A CRC64;
Query Match 10.6%; Score 139; DB 1; Length 231;
Best Local Similarity 27.8%; Pred. No. 3.8e-05;
Matches 65; Conservative 27; Mismatches 96; Indels 46; Gaps 13;
QY 26 LIPELCKQFVHLGVTGTGGGSLKHGNE--IYIAPSGVOKERTOPEDMFCDINEQDIS 83
DB 10 LAANLALPAHLL--VTFWGNVSVADETRQWVVKPSGVEYDVMADVMVVEI-----AS 63
QY 84 GPPASKKLKKSQCCTPLFWNAYTMRG--AGAVITHSKAAVM---ATLLFPQGEFKITHQEM 139
DB 64 GKVVGEKSPSDPTTHALYRRYAEIGGIIVHTRHATWSQAGLDLPA--WGTTHADY 121
QY 140 IKGIRKCTS-----GGYRYDDMLVPIIPIENTPEEKDKERMAHAMNPEYDSCAVLVR 192
DB 122 FYGAIPCTQMTABEINGEYEQ---TGEVLIETFEER-----GRSPAQIP---AVLVH 169
```

QY 193 RHGVYVWG-----ETWEKAKTMCCE-YDYLFDAVSMKMGGLDFTQLPVGENGIV 241
DB 170 SHGPFAWGKNAADAVHNAVLEECAYMGLFSRLA-----PQLPAMQNEILL 215

RESULT 5

```
SGAE_MYCPN      STANDARD;      PRT;      242 AA.
ID  SGAE_MYCPN
AC  P75289;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DE  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Probable sugar isomerase sgae (EC 5.1.-.-).
GN  SGAE OR MPN498 OR MP345.
OS  Mycoplasma pneumoniae.
OC  Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX  NCBI_TaxID=2104;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 29342 / M129;
RC  MEDLINE=97105885; PubMed=8948633;
RA  Himmelsreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA  Herrmann R.;
RT  "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT  pneumoniae.";
RL  Nucleic Acids Res. 24:4420-4449(1996).
CC  -1- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE. PROBABLY
CC  INVOLVED IN A METABOLIC PATHWAY WITH SGAAH AND SGAAU.
CC  -1- COPACTOR: Binds 1 zinc ion per molecule (Potential).
CC  -1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC  subfamily.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; AE000032; AAB95992.1; -.
CC  PIR; S73671; S73671.
CC  InterPro; IPR001303; Aldolase II N.
CC  Pfam; PF00596; Aldolase II; 1.
CC  Isomerase; Zinc; Complete proteome.
CC  METAL 80 80 ZINC (BY SIMILARITY).
CC  METAL 99 99 ZINC (BY SIMILARITY).
CC  METAL 101 101 ZINC (BY SIMILARITY).
CC  METAL 175 175 ZINC (BY SIMILARITY).
CC  SEQUENCE 242 AA; 27092 MW; 3F326B0F364CB8B8 CRC64;
Query Match 10.0%; Score 132; DB 1; Length 242;
Best Local Similarity 27.9%; Pred. No. 0.00017;
Matches 55; Conservative 25; Mismatches 73; Indels 44; Gaps 10;
QY 26 LIPELCKQFVH-----LGWVTGTGGGSLKHGNEIY--IAPSGVOKERTIQEDMFVCD 76
DB 5 MINDLKEQVFTNLLPKYGLVHTWGNVSMIAPNQFFVFKPSGVSYDKQRAQDMVVVD 64
QY 77 I--NEQDISGPPASKKLKKSQCCTPLFWNAYT-MRGAGAVITHSK-AAVMATLLFPQGEF 132
DB 65 LDNNVLDITNG-----LKPSSDPTTHALMYKCPDKAIIVHTSTFTATSPAQADKPIPL 118
QY 133 KITHQEMIKGIRKCTS-----GGYRYDDMLVPIIPIENTPEEKDKERMAHAMNPEYD 185
DB 119 GTTHADNFFGPICTRTALSDEINGAYEHTGLVI-----LEHLKNNQVD 163
QY 186 --SC-AVLNRRHGVYVW 199
DB 164 VNACAILVKEHGSFVW 180
```

RESULT 6

```

SGBE HAEIN          STANDARD;          PRT;          231 AA.
AC P44989;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable sugar isomerase sgbe (EC 5.1.-.-).
GN SGBE OR H1025.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]_TaxID=727;
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uutterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE.
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -!- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32783; AAC22685.1; -.
CC TIGR; H11025; -.
CC InterPro; IPR001303; Aldolase II_N.
CC Pfam; PF00596; Aldolase II; 1.
CC Isomerase; Zinc; Complete proteome.
CC METAL 76 76 ZINC (BY SIMILARITY).
CC FT METAL 95 95 ZINC (BY SIMILARITY).
CC FT METAL 97 97 ZINC (BY SIMILARITY).
CC FT METAL 171 171 ZINC (BY SIMILARITY).
CC SQ SEQUENCE 231 AA; 25980 MW; 9DEB3485E54B10DC7 CRC64;
CC -----
Query Match
Best Local Similarity 9.8%; Score 128.5; DB 1; Length 231;
Matches 50; Conservative 23; Mismatches 74; Indels 33; Gaps 9;
QY 35 YHLGWVTGGGIGSL--KHGNEIYAPSGVQKERIQPEDMFVCDINEQDISPPASKLK 92
DQ 19 HHL--VTFWGNVSAIDREKNLVVTKPSGVVDVNTENDMVVDL-----FTGNIVEGNKK 72
QY 93 KSQCTPLFNAY--TWRGAGAVLTHSK--AAVNATLLFPQEPKITHQEMIKGIRCT--- 147
DQ 73 PSSDPTFHLRYRQPHGIGVHTSRHATIWAQAGLDLIEVGTTHGDFYFTICTPQRM 132
QY 148 -----SGGYRYDDMLVPIIENTPEEKDKERMAHAMEYDPC-AVLVRHGVYVNGE 201
DQ 133 TTKKIKGNY-----ELETGKVIIVETLSRGIE-----PDNIPAVLVHSHGPFAMWK 178

```

RESULT 7

```

ARAD_BACSU          STANDARD;          PRT;          229 AA.
ID ARAD_BACSU

```

```

AC P94525; O05186;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose
DE isomerase).
GN ARAD OR BSU28780.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]_TaxID=1423;
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97237725; PubMed=9084180;
RA Sa-Nogueira I.M.G., Nogueira T.V., Soares S., de Lencastre H.;
RT "The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence,
RT genetic organization and expression.";
RL Microbiology 143:957-969(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124191; PubMed=8969504;
RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emerson P.T., Harwood C.R.;
RT "The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus
RT subtilis chromosome containing genes responsible for stress
RT responses, the utilization of plant cell walls and primary
RT metabolism.";
RL Microbiology 142:3067-3078(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertoletto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriani R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerth I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vaesroeti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RX MEDLINE=99348379; PubMed=10417639;
RA Mota L.J., Tavares P., Sa-Nogueira I.M.G.;
RT "Mode of action of AraR, the key regulator of L-arabinose metabolism
RT in Bacillus subtilis.";
RL Mol. Microbiol. 33:476-489(1999).
CC -!- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-
CC phosphate.

```


CC -|- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -|- PATHWAY: L-arabinose catabolism; third step.
CC -|- INDUCTION: Transcription is repressed by glucose and by the
CC binding of arabinose to the operon promoter. L-arabinose acts as an
CC inducer by inhibiting the binding of arabinose to the DNA, thus
CC allowing expression of the gene.
CC -|- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X89408; CAA61587.1; -
CC EMBL; Z75208; CAA99589.1; -
CC EMBL; Z93118; CAA14838.1; -
CC PIR; E69587; E69587.
CC Subtilisin; BGL1906; arad.
CC InterPro; IPR001303; Aldolase II_N.
CC InterPro; IPR004661; Arad.
CC TIGRfam; TIGR00760; arad; 1.
CC Arabinose catabolism; isomerase; Zinc; Complete proteome.
CC METAL 75 75 ZINC (BY SIMILARITY).
CC METAL 94 94 ZINC (BY SIMILARITY).
CC METAL 96 96 ZINC (BY SIMILARITY).
CC METAL 168 168 ZINC (BY SIMILARITY).
CC CONFLICT 43 48 PSGVEY -> LARSNT (IN REF. 1).
CC CONFLICT 190 190 V -> L (IN REF. 1).
CC SEQUENCE 229 AA; 25686 MW; 43F89C216012E2D3 CRC64;
CC
CC Query Match 9.7%; Score 128; DB 1; Length 229;
CC Best Local Similarity 27.5%; Pred. No. 0.00035;
CC Matches 46; Conservative 25; Mismatches 76; Indels 20; Gaps 8;
CC
CC 40 VTGTGGGIS--LKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKKLKQCT 97
CC Db VTGTGGGIS--LKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKKLKQCT 97
CC 22 VTFTGNVSGIDREKIRIVKPSGVEYSDLTADLVNLNLDGEVVEG-----SLKPSSDT 76
CC
CC 98 PLFNNAY-TMRGAGAVIHTSKAAVMATLLFPFGQEFK---ITHQEMIKGIRKCTGGYYR 153
CC Db PTHVLYLKAFNPGIGIVHTHSQWA--TSWAGSQRDIPLGTTHADYFDSALPCTR---EM 131
CC 154 YDDMLVVPPIENTPEEKDKERMAHMANEYDPDSCAVLVRHGVYVWG 200
CC Db YDELIHHDYELNT--GKVIATFQH--HNYQVQVGVNLVNHGPFVWG 174
CC
CC RESULT 8
CC ARAD_BACST STANDARD; PRT; 228 AA.
CC ID ARAD_BACST STANDARD; PRT; 228 AA.
CC AC Q9S469;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose
CC isomerase).
CC GN ARAD.
CC OS Bacillus stearothermophilus.
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
CC OX NCBI_TaxID=1422;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-T-6;
CC RA Gilead-Gropper S., Shohan Y.;
CC RT "The L-arabinose utilization gene cluster from Bacillus
CC stearothermophilus T-6".
CC RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-

phosphate.
CC -|- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -|- PATHWAY: L-arabinose catabolism; third step.
CC -|- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF160811; AAD45716.1; -
CC InterPro; IPR001303; Aldolase II_N.
CC Pfam; PF00596; Aldolase II; 1.
CC Arabinose catabolism; isomerase; Zinc.
CC METAL 75 75 ZINC (BY SIMILARITY).
CC METAL 94 94 ZINC (BY SIMILARITY).
CC METAL 96 96 ZINC (BY SIMILARITY).
CC METAL 168 168 ZINC (BY SIMILARITY).
CC SEQUENCE 228 AA; 25227 MW; 66EEC1746D950C84 CRC64;
CC
CC Query Match 9.5%; Score 125.5; DB 1; Length 228;
CC Best Local Similarity 24.7%; Pred. No. 0.00059;
CC Matches 47; Conservative 21; Mismatches 69; Indels 53; Gaps 7;
CC
CC 33 QFVHLGWVTGGISLKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKKLK 92
CC Db QFVHLGWVTGGISLKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPPASKKLK 92
CC 18 QYRLVFTTGNVSGIDREKIRIVKPSGVEYSDLTADLVNLNLDGEVVEG-----DLK 71
CC
CC 93 KSQCTPLFNNAY-TMRGAGAVIHTSKAAVMATLLFPFGQEFKITHQEMIKGIRKCT--- 147
CC Db PSSDTTHLWLYKQFGIGIVHTHSQWA--TSWAGSQRDIPLGTTHADYFDSALPCTR 131
CC 148 -----SGGY-----RYDDMLVVPPIENTPEEKDKERMAHMANEYDPDSCAVL 191
CC Db TNEEIQAYELETGKVTETFRFLDPLQMB-----GVLV 165
CC 192 RRGVYVWGE 201
CC Db HGHGPFVWGE 175
CC
CC RESULT 9
CC ARAD_BACST STANDARD; PRT; 231 AA.
CC AC Q9KBO4;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose
CC isomerase).
CC GN ARAD OR BH1871.
CC OS Bacillus halodurans.
CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC OX NCBI_TaxID=86665;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=C-125 / JCM 9153;
CC RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
CC Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
CC Horikoshi K.;
CC RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
CC halodurans and genomic sequence comparison with Bacillus subtilis".
CC Nucleic Acids Res. 28:4317-4331(2000).
CC -|- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-
CC phosphate.
CC -|- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -|- PATHWAY: L-arabinose catabolism; third step.
CC -|- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca

CC subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF001513; BAB05590.1; -.
 CC PIR; G83883; G83883.
 CC InterPro; IPR001303; Aldolase II_N.
 CC Pfam; PF00596; Aldolase II; 1.
 CC Arabinose catabolism; Isomerase; Zinc; Complete proteome.
 KW METAL 75 75 ZINC (BY SIMILARITY).
 FT METAL 94 94 ZINC (BY SIMILARITY).
 FT METAL 96 96 ZINC (BY SIMILARITY).
 FT METAL 168 168 ZINC (BY SIMILARITY).
 FT METAL 231 AA; 25991 MW; 6CA570D2A6B4DB97 CRC64;
 SQ SEQUENCE 231 AA; 25991 MW; 6CA570D2A6B4DB97 CRC64;
 Query Match 9.4%; Score 124; DB 1; Length 231;
 Best Local Similarity 28.7%; Pred. No. 0.00081;
 Matches 51; Conservative 20; Mismatches 67; Indels 40; Gaps 10;
 CC
 CC 40 VTGTGGIS--LKHGNEIYIAPSGVQKRIQPEDMFVCDINEODISGPPASKKLKKSQCT 97
 CC 22 VFTTGNVSGIDREKGLVIVKPSGVYFEMKSKDMVVLDGNGIVEG---DLKPSST 76
 CC
 CC 98 PLPMNAY-TMRGAGAVIHTSKAAVMATL-LFGQEQ---FKITHQEMIKGIRKCTS--- 148
 CC 77 PTHLALYRAPDKVGGIVHTS---VWATAWAQAKGEPAYGTTTHADYFHTGTPCTPMT 133
 CC
 CC 149 ---GGYRYDDMLVPIIENTPPEKDKERMAHAMNEVPDSC-AVLVRHGVYVNGE 201
 CC 134 TEILGDYKETGNVIVETFRN---KD-----PMSIPGLVHSHAFVFWGK 175
 CC
 CC RESULT 10
 CC SGAE_ECOLI STANDARD; PRT; 228 AA.
 CC ID SGAE_ECOLI
 CC AC P39306;
 CC DT 01-FEB-1995 (Rel. 31, Created)
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Probable sugar isomerase sgae (EC 5.1.-.-).
 CC GN SGAE OR B4198.
 CC OS Escherichia coli.
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC OC Enterobacteriaceae; Escherichia.
 CC OX NCBI_TaxID=562;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=K12 / MG1655;
 CC RX MEDLINE=95334362; PubMed=7610040;
 CC RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 CC Blattner F.R.;
 CC RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 CC RT region from 92.8 through 100 minutes."
 CC RL Nucleic Acids Res. 23:2105-2119(1995).
 CC RN [2]
 CC RP DISCUSSION OF SEQUENCE.
 CC RA Reizer J., Charbit A., Reizer A., Sailer M.H. Jr.;
 CC RT "Novel phosphotransferases system genes revealed by bacterial genome
 CC RT analysis: operons encoding homologues of sugar-specific permease
 CC RT domains of the phosphotransferase system and pentose catabolic
 CC RT enzymes."
 CC RL Genome Sci. Technol. 1:53-75(1996).
 CC CC -1- FUNCTION: PROBABLE PENTULOSE-5-PHOSPHATE-4-EPIMERASE. PROBABLY
 CC INVOLVED IN A METABOLIC PATHWAY WITH SGAE AND SGAT.
 CC CC -2- COFACTOR: Binds 1 zinc ion per molecule (Potential).
 CC CC -3- SIMILARITY: Belongs to the aldolase class II family. AraD/fucA
 CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14003; AAA97094.1; -.
 CC PIR; S56423; S56423.
 CC EcoGene; EG12498; sgae.
 CC InterPro; IPR001303; Aldolase II_N.
 CC Pfam; PF00596; Aldolase II; 1.
 CC Isomerase; Zinc; Complete proteome.
 KW METAL 74 74 ZINC (BY SIMILARITY).
 FT METAL 93 93 ZINC (BY SIMILARITY).
 FT METAL 95 95 ZINC (BY SIMILARITY).
 FT METAL 167 167 ZINC (BY SIMILARITY).
 FT METAL 228 AA; 25278 MW; 3E96E7E0261E36B6 CRC64;
 SQ SEQUENCE 228 AA; 25278 MW; 3E96E7E0261E36B6 CRC64;
 Query Match 9.1%; Score 120; DB 1; Length 228;
 Best Local Similarity 25.8%; Pred. No. 0.0018;
 Matches 50; Conservative 24; Mismatches 74; Indels 46; Gaps 9;
 CC
 CC 21 EHRFLIPELCKQFVHLGWVTGGISL--KHGNEIYIAPSGVQKRIQPEDMFVCDIN 78
 CC 14 ELPRY-----GLVTTGNVNSAIDRGLWIKPSGVAYETMKAADMVVVDMS 61
 CC
 CC 79 EQDISGPPASKKLKKSQCTPLFMNAY-TMRGAGAVIHTSKAA---VMATLLFPFGQEFKI 134
 CC 62 GKVEG-----FYRPSSTATHLELYRRYPSLGGIVHTSTATAWAQAGLAIPA--LGT 114
 CC
 CC 135 THQEMIKGIRKCTSG-----GYRYDDMLVPIIENTPPEKDKERMAHAMNEVPDSC 187
 CC 115 THADYFHTGTPCTPCTGLSBEVQGEYELNTGKV--IETLGNAPL-----HTP 160
 CC
 CC 188 AVLVRHGVYVNGE 201
 CC 161 GIVVYQHGPFAWGK 174
 CC
 CC RESULT 11
 CC ARAD_SALTY STANDARD; PRT; 231 AA.
 CC ID ARAD_SALTY
 CC AC P06130;
 CC DT 01-JAN-1988 (Rel. 06, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose
 CC DE isomerase).
 CC GN ARAD OR STM0101.
 CC OS Salmonella typhimurium.
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC OC Enterobacteriaceae; Salmonella.
 CC OX NCBI_TaxID=602;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=LT2;
 CC RX MEDLINE=85222046; PubMed=3891514;
 CC RA Lin H.-C., Lei S.-P., Studnicka G., Wilcox G.;
 CC RT "The arabid operon of Salmonella typhimurium LT2. III. Nucleotide
 CC RT sequence of arab and its flanking regions, and primary structure of
 CC RT its product, L-ribulose-5-phosphate 4-epimerase."
 CC RL Gene 34:129-134(1985).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 CC RX MEDLINE=21534948; PubMed=11677609;
 CC RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 CC Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 CC Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856 (2001).
 CC 1- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-
 CC phosphate.
 CC 1- COFACTOR: Binds 1 zinc ion per molecule (Potential).
 CC 1- PATHWAY: L-arabinose catabolism; third step.
 CC 1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
 CC subfamily.
 CC 1- CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 202.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M11047; AAA27025.1; ALT FRAME.
 CC EMBL: AE008698; AAL19065.1; -.
 CC PIR: A24986; ISEB4T.
 CC StyGene; SG10015; arad.
 CC InterPro; IPR001303; Aldolase II_N.
 CC InterPro; IPR004661; Arad.
 CC Pfam; PF00596; Aldolase II; 1.
 CC TIGRFAMs; TIGR00760; arad; 1.
 CC Arabinose catabolism; Isomerase; Zinc; Complete proteome.
 FT METAL 76 76 ZINC (BY SIMILARITY).
 FT METAL 95 95 ZINC (BY SIMILARITY).
 FT METAL 97 97 ZINC (BY SIMILARITY).
 FT METAL 171 171 ZINC (BY SIMILARITY).
 SQ SEQUENCE 231 AA; 25531 MW; DA473505739284F9 CRC64;
 Query Match 8.6%; Score 113.5; DB 1; Length 231;
 Best Local Similarity 26.4%; Pred. No. 0.0068;
 Matches 47; Conservative 20; Mismatches 84; Indels 27; Gaps 7;
 QY 40 VTGTGGGSL-KHGNEIYAPSGVKERIQEDMFVCDINEQDISPPASKLKKSQCT 97
 Db 22 VLTGWNVSVDREGVLVIRPSGVDYSVMTADNVVSL- - -SGEVGHHKPSD 77
 QY 98 PLFMNAY-TMRGAGAVITHSK-AAVMATLIPPGQEFKITHQEMIKGRKCTS 148
 Db 78 PTHRLLYQAFPTIGGVTHSRHATIAQAQGIPTATGTTHADYFGYPTICTRKMTAEAI 137
 QY 149 GGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVLRHGVVWGTEWEKA 206
 Db 138 NGEYEW- - -TGNVIVTFEKQGI- - - - -DAAQMPGVLVHSHGFPAMGNAEDA 183
 RESULT 12
 FUCA_HAEIN STANDARD; PRT; 216 AA.
 AC P44777;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE L-fuculose phosphate aldolase (EC 4.1.2.17) (L-fuculose-1-phosphate
 DE aldolase).
 GN FUCA OR H10611.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 ON NCBI_TaxId=727;
 RX STRAIN-Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Spriggs T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 RT Rd.";
 RL Science 269:496-512 (1995).
 CC 1- CATALYTIC ACTIVITY: L-fuculose 1-phosphate = glycerone phosphate +
 CC (S)-lactaldehyde.
 CC 1- COFACTOR: Binds 1 zinc ion per molecule (Potential).
 CC 1- PATHWAY: Fucose metabolism; third step.
 CC 1- SUBUNIT: Homotetramer (By similarity).
 CC 1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca
 CC subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U32743; AAC22270.1; -.
 CC PIR: C64081; C64081.
 CC HSP; P11550; 2FUA.
 CC TIGR; H10611; -.
 CC InterPro; IPR001303; Aldolase II_N.
 CC InterPro; IPR004782; Fuca.
 CC Pfam; PF00596; Aldolase II; 1.
 CC TIGRFAMs; TIGR01086; fuca; 1.
 CC Lyase; Fucose metabolism; Zinc; Complete proteome.
 FT METAL 73 73 ZINC (BY SIMILARITY).
 FT METAL 92 92 ZINC (BY SIMILARITY).
 FT METAL 94 94 ZINC (BY SIMILARITY).
 FT METAL 155 155 ZINC (BY SIMILARITY).
 SQ SEQUENCE 216 AA; 23944 MW; 4DBEE21ED8EA6C6 CRC64;
 Query Match 8.4%; Score 110; DB 1; Length 216;
 Best Local Similarity 23.3%; Pred. No. 0.013;
 Matches 48; Conservative 34; Mismatches 96; Indels 28; Gaps 6;
 QY 27 IPELCKQFHLGWTGTGGGSLKHGNEIYAPSGVKERIQEDMFVCDINEQDISGPP 86
 Db 10 IIDTCLMTKLGNGQTAGNVSVRYKQGMILITPTGMPYHLMKNTENIVVDGN- - -GKH 64
 QY 87 ASKKLKKSQCTPLFMNAYTMRGAGAVITHSKAAVMATL- - -FPQGFKITHQEMIKGI 143
 Db 65 EENKLPSEWQFHLVSVHTPREANAVVNHSHHCAGLSILEKPIAIHVV- - - - - 115
 QY 144 RKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMNEYPDSCAVLRHGVVWGTEW 203
 Db 116 - - -AVSG- - -TDHIPCPVYA- - -TFGSHKLSYVATGIKE- - -SKALLAHHLITCGENL 164
 QY 204 EKATMCECYDYLFDIAVSMKKMGILD 229
 Db 165 DRALMLAQEVEVLASVWLKLLSTGLE 190
 RESULT 13
 ARAD_ECOLI STANDARD; PRT; 231 AA.
 ID ARAD_ECOLI
 AC P08203;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose
 DE isomerase).
 GN ARAD OR B0061.

OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC STRAIN=B;
RX MEDLINE=87163495; PubMed=3549454;
RA Lee N., Gleisow W., Martin R., Hamilton E., Fowler A.;
RT "The organization of the araBAD operon of Escherichia coli.";
RL Gene 47:231-244(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=91083835; PubMed=2261080;
RA Chen H., Sun Y., Stark T., Beattie W., Moses R.E.;
RT "Nucleotide sequence and deletion analysis of the polB gene of
RT Escherichia coli";
RL DNA Cell Biol. 9:631-635(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91067495; PubMed=2251150;
RA Mineno J., Fukui H., Ishino Y., Kato I., Shinagawa H.;
RT "Nucleotide sequence of the araD gene of Escherichia coli K12
RT encoding the L-ribulose 5-phosphate 4-epimerase";
RL Nucleic Acids Res. 18:6722-6722(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12; MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaser J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [6]
RP SEQUENCE OF 222-231 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91017565; PubMed=2217198;
RA Bonner C.A., Hays S., McEntee K., Goodman M.F.;
RT "DNA polymerase II is encoded by the DNA damage-inducible dinA gene
RT of Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7663-7667(1990).
RN [7]
RP SEQUENCE OF 158-231 FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=91238699; PubMed=2034216;
RA Iwasaki H., Ishino Y., Toh H., Nakata A., Shinagawa H.;
RT "Escherichia coli DNA polymerase II is homologous to alpha-like DNA
RT polymerases";
RL Mol. Gen. Genet. 226:24-33(1991).
RN [8]
CC -!- CATALYTIC ACTIVITY: L-ribulose 5-phosphate = D-xylulose 5-
CC phosphate.
CC -!- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -!- PATHWAY: L-arabinose catabolism; third step.
CC -!- SIMILARITY: Belongs to the aldolase class II family. AraD/fucA
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15263; AAA23464.1; -;
DR EMBL; M35371; -; NOT ANNOTATED_CDS.
DR EMBL; M62646; AAA24405.1; -;
DR EMBL; D10483; BAB96630.1; -;
DR EMBL; AE000116; AAC73172.1; -;
DR EMBL; M37727; AAA23683.1; -;
DR EMBL; M38283; AAA63763.1; -;
DR EMBL; X56048; CAA39519.1; -;
DR PIR; E64727; ISECP4.
DR PDB; 1JDI; 23-JAN-02.
DR PDB; 1KOW; 28-JAN-03.
DR ECO2DBASE; G028.1; 6TH EDITION.
DR EcoGene; EG10055; araD.
DR InterPro; IPR001303; Aldolase_II_N.
DR InterPro; IPR004661; AraD.
DR Pfam; PF00596; Aldolase_II; 1.
DR TIGRFAMs; TIGR00760; araD; 1.
KW Arabinose catabolism; Isomerase; Zinc; Complete proteome;
KW 3D-structure.
FT METAL 76 76 ZINC (BY SIMILARITY).
FT METAL 95 95 ZINC (BY SIMILARITY).
FT METAL 97 97 ZINC (BY SIMILARITY).
FT METAL 171 171 ZINC (BY SIMILARITY).
FT VARIANT 50 50 V -> I.
FT VARIANT 70 70 T -> A.
FT VARIANT 216 216 D -> N.
SQ SEQUENCE 231 AA; 25519 MW; 1753F75958332163 CRC64;
Query Match 8.3%; Score 109.5; DB 1; Length 231;
Best Local Similarity 26.1%; Pred. No. 0.015;
Matches 47; Conservative 19; Mismatches 83; Indels 31; Gaps 8;
QY 40 VTGTGGGIGSL--KHGNEIYIAPSGVQKRIQEDMFVCDINEQDISGPPASKKLKKSQCT 97
DB 22 VTLTWGNVSAVDRERGVRGKPSGVYDYSVMTADDMVVVSIE---TGEVVEGTKKPSST 77
QY 98 PLFMNAY--TMRGAGAVIHTHSAKAAVWATLLPQQEFK---ITHQEMIKGIKICTS----- 148
DB 78 FTHRLLYQAPPSIGGIVHSHRHAIWAQ--AGQSIPTGTHADYFYGTICTRKMTDA 135
QY 149 --GGYRYDDMLVPIIPIENTPEEKDKERMAHAMNEYDSCAVLVRHGVVYVWGETWEKA 206
DB 136 EINGEYEW---TGNVIVETFEKQGI-----DAAQMPGLVHSHGPFAGKQAEDA 183
RESULT 14
FUCA_ECOLI STANDARD; PRT; 215 AA.
AC P11550;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE L-fuculose phosphate aldolase (EC 4.1.2.17) (L-fuculose-1-phosphate
DE aldolase).
GN FUCA OR FUCC OR PRD OR B2800 OR SF2814 OR S3009.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RN SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=K12;
RC MEDLINE=89315234; PubMed=2664711;
RX Lu Z., Lin B.C.C.;
RA "The nucleotide sequence of Escherichia coli genes for L-fucose
RT dissimilation";
RL Nucleic Acids Res. 17:4883-4884(1989).
RN [2]

RN SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=90036697; PubMed=2553671;
RA Chen Y.M., Lu Z., Lin E.C.C.;
RT "Constitutive activation of the fucAO operon and silencing of the
RT divergently transcribed fucIK operon by an ISS element in
RT Escherichia coli mutants selected for growth on L-1,2-propanediol.";
RL J. Bacteriol. 171:6097-6105(1989).
RN [3]
RN SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RN SEQUENCE OF 108-215 FROM N.A.
RP SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=89291720; PubMed=2661535;
RA Conway T., Ingram L.O.;
RT "Similarity of Escherichia coli propanediol oxidoreductase (fucO
RT product) and an unusual alcohol dehydrogenase from Zymomonas mobilis
RT and Saccharomyces cerevisiae.";
RL J. Bacteriol. 171:3754-3759(1989).
RN [5]
RN SEQUENCE FROM N.A.
RP SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [6]
RN SEQUENCE FROM N.A.
RP SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.R., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
RN [7]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS), AND CATALYTIC MECHANISM.
RP SPECIES=E.coli;
RX MEDLINE=96256522; PubMed=8676381;
RA Dreyer M.K., Schulz G.E.;
RT "Catalytic mechanism of the metal-dependent fuculose aldolase from
RT Escherichia coli as derived from the structure.";
RL J. Mol. Biol. 259:458-466(1996).
RN [8]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RP SPECIES=E.coli;
RX MEDLINE=96256522; PubMed=8676381;
RA Dreyer M.K., Schulz G.E.;
RT "Refined high-resolution structure of the metal-ion dependent L-
RT fuculose-1-phosphate aldolase (class II) from Escherichia coli.";
RL Acta Crystallogr. D 52:1082-1091(1996).
RN [9]
RN X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS), AND MUTAGENESIS.
RP SPECIES=E.coli;
RX MEDLINE=20281325; PubMed=10821675;
RA Joergers A.C., Grosse C., Fessner W.D., Schulz G.E.;
RT "Catalytic action of fuculose 1-phosphate aldolase (class II) as
RT derived from structure-directed mutagenesis.";
RL Biochemistry 39:6033-6041(2000).

RN X-RAY CRYSTALLOGRAPHY OF MUTANTS.
RP SPECIES=E.coli;
RX MEDLINE=20510153; PubMed=11054289;
RA Joergers A.C., Mueller-Dieckmann C., Schulz G.E.;
RT "Structures of L-fuculose-1-phosphate aldolase mutants outlining
RT motions during catalysis.";
RL J. Mol. Biol. 303:531-543(2000).
RN [10]
RN CATALYTIC ACTIVITY: L-fuculose 1-phosphate = glycerone phosphate +
RN (S)-lactaldehyde.
CC - COFACTOR: Binds 1 zinc ion per molecule.
CC - PATHWAY: Fucose metabolism; third step.
CC - SUBUNIT: Homotetramer.
CC - SIMILARITY: Belongs to the aldolase class II family. Arab/fucA
CC subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M31059; AAA23823.1; -
CC EMBL; X15025; CAA33125.1; -
CC EMBL; U29581; AAB40450.1; -
CC EMBL; AE000363; AAC75842.1; -
CC EMBL; M27177; -; NOT ANNOTATED_CDS.
CC EMBL; AE015295; AAN44302.1; -
CC EMBL; AE016987; AAP18127.1; -
CC PIR; B33495; ADECFP.
CC PDB; 1FUA; 14-OCT-96.
CC PDB; 2FUA; 14-OCT-96.
CC PDB; 3FUA; 14-OCT-96.
CC PDB; 4FUA; 14-OCT-96.
CC PDB; 1E46; 18-JUL-03.
CC PDB; 1E47; 18-JUL-03.
CC PDB; 1E48; 18-JUL-03.
CC PDB; 1E49; 18-JUL-03.
CC PDB; 1E4A; 18-JUL-03.
CC PDB; 1E4B; 18-JUL-03.
CC PDB; 1E4C; 18-JUL-03.
CC PDB; 1D2V; 21-FEB-02.
CC PDB; 1D2U; 20-JUL-00.
CC PDB; 1D2W; 21-FEB-02.
CC PDB; 1D2X; 21-FEB-02.
CC PDB; 1D2Y; 24-JUN-03.
CC PDB; 1D2Z; 21-FEB-02.
CC EcoGene; EGI0348; fucA.
CC InterPro; IPR001303; Aldolase_II_N.
CC Pfam; PF00596; Aldolase_II; 1.
CC TIGRfam; TIGR01086; fucA; 1.
KW Lyase; Fucose metabolism; Zinc; 3D-structure; Complete proteome.
FT METAL 73 73 ZINC.
FT METAL 92 92 ZINC.
FT METAL 94 94 ZINC.
FT METAL 155 155 ZINC.
FT HELIX 3 19
FT TURN 20 21
FT STRAND 24 26
FT STRAND 29 34
FT TURN 35 36
FT STRAND 37 40
FT STRAND 42 42
FT HELIX 47 49
FT STRAND 52 54
FT TURN 56 58
FT STRAND 60 61
FT TURN 64 64
FT STRAND 66 66
FT TURN 70 70
FT STRAND 70 70

FT TURN 72 73
FT HELIX 74 83
FT TURN 85 86
FT STRAND 89 93
FT HELIX 96 104
FT TURN 105 105
FT STRAND 108 108
FT HELIX 113 118
FT TURN 119 119
FT STRAND 123 123
FT STRAND 125 126
FT TURN 131 132
FT TURN 134 143
FT STRAND 149 152
FT TURN 153 155
FT STRAND 156 161
FT HELIX 164 185
FT TURN 186 187
FT HELIX 196 209
SQ SEQUENCE 215 AA; 23775 MW; BA9897E13ABE4A22 CRC64;

Query Match 8.2%; Score 108.5; DB 1; Length 215;
Best Local Similarity 23.9%; Pred. No. 0.017;
Matches 52; Conservative 37; Mismatches 86; Indels 43; Gaps 11;

QY 27 IPELCKQFYHLGWVTGGTGGISLKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGPP 86
DB 10 IDTTCLEMTFLGNQTAGNSVRYQDGMILITPTGIPYKLTESHIVFDGNGKHEEG-- 67
QY 87 ASKKUKKSQCTLFNNAYTMR-GAGAVIHTHSKAAVMATLL---PPQBEFKITHQEMIKG 142
DB 68 ---KLPSSEWR-FHMAAYQSRPDANAVVHNHVAHVCTAVSILNRSIPAIHYMI----- 115
QY 143 IRKTSGGYRYDDMLVPIENTPEEKDKERMAHAMEYDPSCAVLVRRHGVVWGRT 202
DB 116 ---AAAGG---NSIPCAVYA-TFTRELSERVALALK---NRKATLQHGHLACEVN 163
QY 203 WEKAKTMCBCYD-----YLFDAVSMKMGGLDPTQLPV 235
DB 164 LEKALWLAHEVEVLAQLYLTTLAIT-----DP--VPV 193

RESULT 15
YGBL HAEIN STANDARD; PRT; 210 AA.
AC Q57199; O05040;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical aldolase class II protein H11012.
GN H11012.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -1- COFACTOR: Binds 1 zinc ion per molecule (Potential).
CC -1- SIMILARITY: Belongs to the aldolase class II family. Arad/fuca

CC subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32782; AAC22673.1; -;
DR TIGR; B64108; B64108.
DR TIGR; H11012; -;
DR InterPro; IPR001303; Aldolase II_N.
DR Pfam; PF00596; Aldolase II; 1.
KW Hypothetical protein; zinc; Complete proteome.
FT METAL 74 74 ZINC (BY SIMILARITY).
FT METAL 93 93 ZINC (BY SIMILARITY).
FT METAL 95 95 ZINC (BY SIMILARITY).
FT METAL 160 160 ZINC (BY SIMILARITY).
SQ SEQUENCE 210 AA; 23242 MW; 4800CA987DBDF01F CRC64;

Query Match 7.8%; Score 102.5; DB 1; Length 210;
Best Local Similarity 22.3%; Pred. No. 0.058;
Matches 41; Conservative 26; Mismatches 82; Indels 35; Gaps 6;

QY 26 LIPELCKQFYHLGWVTGGTGGISLKHGNEIYIAPSGVQKRIQPEDMFVCDINEQDISG 84
DB 9 LMVQLGRSFYERGIYVGGAGNLSVFLDDNRVLVPTGTSSLGRLSVERLSVLDMEGNLLGG 68
QY 85 PPASKKL-----KKSQCTPLFNNAYTMRGAGAVIHTHSKAAVMATLLFPQBEFKITH 136
DB 69 DKPSKEAVFHLAMYKKNPECK-----AIVHLHSTYLTALSCL-----DN 107
QY 137 QEMIKGIRKCTSGGYRYDDMLVPIENTPEEKDKERMAHAMEYDPSCAVLVRRHGV 196
DB 108 LDPNNAIPPTPYVNRVGMQVPIYYR--PGSPKIAELS---NRALTGKAFLLANHGV 162
QY 197 YVWG 200
DB 163 VTG 166

Search completed: June 1, 2004, 13:33:52
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:30:06 ; Search time 20 seconds
(without alignments)
1159.108 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQGCCRPQGAQDK.....SMKMGELDPTQLPVGENGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507	38.5	244	2 S57042	hypothetical prote
2	433.5	32.9	284	2 T27523	hypothetical prote
3	274.5	20.8	192	2 T39191	conserved hypothet
4	219.5	16.7	209	2 A69864	conserved hypothet
5	198.5	14.3	208	1 H70469	L-fuculose-phospha
6	164	12.5	205	2 D83436	probable sugar ald
7	153	11.6	218	2 B82587	conserved hypothet
8	152	11.5	181	1 A64477	L-fuculose-phospha
9	148.5	11.3	254	1 A72396	L-fuculose-phospha
10	140	10.6	244	2 AC3533	L-fuculose-phospha
11	139	10.6	231	2 S47804	L-ribulose-phospha
12	138.5	10.5	191	1 C69054	L-fuculose-phospha
13	132	10.0	231	2 A10977	probable sugar iso
14	132	10.0	242	2 S73671	L-ribulose-phospha
15	132	10.0	243	2 G90586	sugar isomerase sg
16	130	9.9	227	2 A95238	hypothetical prote
17	128.5	9.8	243	2 H64108	L-ribulose-phospha
18	128	9.7	229	2 E59587	L-ribulose-phospha
19	127	9.6	211	2 H84275	fuculose-1-phospha
20	126	9.6	230	2 B82484	sugar isomerase Sg
21	125	9.5	234	2 B98102	L-ribulose-phospha
22	124	9.4	231	2 G83883	L-ribulose-5-phosp
23	123.5	9.4	215	1 E72546	L-fuculose-phospha
24	122.5	9.3	231	2 A10271	L-ribulose-phospha
25	122	9.3	228	2 F91275	probable epimerase
26	122	9.3	228	2 F86116	probable epimerase
27	121	9.2	228	2 A01052	probable class II
28	120	9.1	228	2 S56423	L-ribulose-phospha
29	115.5	8.8	231	2 A80404	L-ribulose-phospha

30	114	8.7	189	1 E71241	L-fuculose-phospha
31	113.5	8.6	248	1 ISEB4T	L-ribulose-phospha
32	111.5	8.5	212	2 C95253	L-fuculose-phospha
33	111.5	8.5	215	2 AH0862	fuculose-1-phospha
34	111.5	8.5	217	2 A98118	L-fuculose-phospha
35	110.5	8.4	231	2 AB0515	L-ribulose-5-phosp
36	110.5	8.4	233	2 B97065	ribulose-5-phospha
37	110	8.4	216	1 C64081	L-fuculose-phospha
38	109.5	8.3	231	1 ISECP4	L-ribulose-5-phosp
39	109.5	8.3	231	2 A90637	L-ribulose-5-phosp
40	109.5	8.3	231	2 A85488	L-ribulose-5-phosp
41	108.5	8.2	215	1 ADECFP	L-fuculose-phospha
42	108.5	8.2	215	2 D91086	L-fuculose-1-phosp
43	108.5	8.2	215	2 F85931	L-fuculose-1-phosp
44	102.5	7.8	210	1 B64108	L-fuculose-phospha
45	99.5	7.6	224	2 AF3646	1-fuculose phospha

ALIGNMENTS

RESULT 1

S57042
hypothetical protein YUR024c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein J1545; hypothetical protein YUR83.18
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence revision 08-Sep-1995 #text_change 19-Apr-2002
C;Accession: S57042; S57039; S55213; S60503; S61120
R;Zagulska, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herberich
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57040
A;Accession: S57042
A;Molecule type: DNA
A;Residues: 1-244 <ZAG>
A;Cross-references: EMBL:Z49524; NID:G1015662; PIDN:CAA89549.1; PID:G1015663; MIPS:YUR024c
R;de Haan, M.; Grivell, L.A.; Smits, P.H.M.; Grivell, L.A.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56771
A;Accession: S57039
A;Molecule type: DNA
A;Residues: 44-244 <ZAG>
A;Cross-references: EMBL:Z49524; MIPS:YUR024c
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A;Reference number: S55183
A;Accession: S55213
A;Molecule type: DNA
A;Residues: 44-244 <DEH>
A;Cross-references: EMBL:X87611; NID:G854567; PIDN:CAA60947.1; PID:G854598
R;Zagulska, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herberich
Yeast 11, 1179-1186, 1995
A;Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading fr
A;Reference number: S60503; MUID:96109930; PMID:8619316
A;Accession: S60503
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-244 <ZAF>
A;Cross-references: EMBL:X87297; NID:G1129159; PIDN:CAA60719.1; PID:e1833285; PID:G1129166
C;Genetics:
A;Cross-references: SGD:S0003785
A;Map position: 10R

Query Match 38.5%; Score 507; DB 2; Length 244;

Best Local Similarity 46.0%; Pred. No. 4, 7e-38;

Matches 108; Conservative 36; Mismatches 67; Indels 24; Gaps 7;

Qy 14 PCGAQDKHEPRFLPELCKQFYHLGWVTGTGGGSLK--HGNEIYIAPSGVQKEKRIOPED 71

Db 13 PC-----HPANLICLTCKQFFHNWCTGTGGGSIKDPNTNYVYLLAPSGVQKEKMPED 66

Qy 72 MFVCDINEQDISPPASKKLKKSQCTELFNWYTMRCGAVIHTHSAAYWATLLFPQGE 131

Db 67 LFVMDAQTLLEYLRSP--KLYKPSACTFLCYQKQKAGAIHTHSONAVICSLLF-GDE 123

Query Match 20.8%; Score 274.5; DB 2; Length 192;
Best Local Similarity 34.6%; Pred. No. 2.9e-17;

RESULT 5

H70469

L-fuculose-phosphate aldolase homolog - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: H70469
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oviatt, D.

Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: H70469
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-208 <AQF>
A:Cross-references: GB:AE000766; GB:AE000657; NID:g2984216; PIDN:AAC07751.1; PID:g298422

A:Experimental source: strain VF5
C:Genetics:
A:Gene: fucA2
C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match 14.3%; Score 188.5; DB 1; Length 208;
Best Local Similarity 24.6%; Pred. No. 1.7e-09;
Matches 52; Conservative 39; Mismatches 93; Indels 27; Gaps 5;

QY 29 ELCKQF-----YHLGWTGTGGGSLKHGNE-IYIAPSGVQKRIQPEDMF 73

Db 4 ELFKFSEKVERIEAGRIILHSRGWVPATSGNISAKVSEYIATASGKHGKLTPELIL 63

QY 74 VQDINEQDI-SGPPASKLKKSOCTPLFNAYTMRGAGAVIHTHSAANWATLLFPQEF 132

Db 64 LIDYGRPVGGKPSAETLLHTTVYKLFPE-----VNAVHTHSPNATVISIVEKKDFV 117

QY 133 KTHQEMIKGIRKCTSGGYRYDDMLVPIENTPEKDKERMAHAMNEYPDSCAVLVR 192

Db 118 ELEDYELLKAF-----PDITHREVKIKIPFPNEQNIPLAKEVNYFKTSEDKYGLIR 172

QY 193 RHGVYVWGETWEKAKTMCCECYDLPDIAYSM 223

Db 173 GHGLYWGSRMEEALHTHEALEFIFCECLKL 203

RESULT 6

D83436
probable sugar aldolase PAL683 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83436
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bohn, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, S.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83436

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-205 <STO>

A:Cross-references: GB:AE004595; GB:AE004091; NID:g9947644; PIDN:AAG05072.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PAL683

Query Match 12.5%; Score 164; DB 2; Length 205;
Best Local Similarity 25.9%; Pred. No. 2.7e-07;
Matches 51; Conservative 28; Mismatches 98; Indels 20; Gaps 5;

QY 24 RFLPELCKQFYHLGWTGTGGGSLKHGNE-IYIAPSGVQKRIQPEDMFVCDINEQDI 82

Db 17 RFL-----YRGKNSPATSSNYSARLDQRALLTVSGKHGKGLGDFDVLA-----TDL 63

QY 83 SGPPASKLKKSOCTPLFNAYTMRGAGAVIHTHSAANWATLLFPQEFKITHQEMIK 141

Db 64 AGNSLEPGKPSAETLLHTTQYANPAIGAVALHTHSAANWATLLFPQEFKITHQEMIK 123

QY 142 GIRKCTSGGYRYDDMLVPIENTPEKDKERMAHAMNEYPDSCAVLVRHGVYVWGE 201
Db 124 AF-----AGVTTHQGVPEPIFDNDQDIARLASRVQVPLEAHPHCPGILRHGHLYTWGA 178
QY 202 TWEKAKTMCCECYDLYFD 218
Db 179 RMSDALRQVEAFELFE 195

RESULT 7

E82587
conserved hypothetical protein XF2209 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: E82587
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82587
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-218 <SIM>
A:Cross-references: GB:AE004033; GB:AE003849; NID:g9107342; PIDN:AAF85008.1; GSPDB:GN001

A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neco, E.; Docena, C.; El-Dorri, H.; Pacincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, R.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2209

Query Match 11.6%; Score 153; DB 2; Length 218;
Best Local Similarity 23.2%; Pred. No. 2.8e-06;
Matches 48; Conservative 34; Mismatches 91; Indels 34; Gaps 7;

QY 33 QFVHLGWVTGGGSLK-HGNEIYIAPSGVQKRIQPEDMFVCDINEQDISGP--PASK 89

Db 29 EFAQAGTTPATSSNFSHRLDEHVAITVSGRDKRCLTEEDIMAVDLGNVAGHPHTPSAE 88

QY 90 KKKKSCTPLFNAYTMRGAGAVIHTHSAANWATLLFPQO-EFKLTHQEMIKGIRKCTS 148

Db 89 TLLHTLYRRFPE-----IGCVLHTSHSTQTVASRYAGAGHISKDYELLKAFE---- 138

QY 149 GGYRYDDMLVPIENTPEKDL-----KERMAHAMNEYPDSCAVLVRHGVYVW 199

Db 139 -GSHTHETILDVFPVFNCTQNNILAAQVDTLLDKQRM-----WGLINGHGMVYTW 187

QY 200 GETWEKAKTMCCECYDLYFDIAVSMKKM 226

Db 188 GNTLADARRHLEALEFLLHCELNLKL 214

RESULT 8

A64477

L-fuculose-phosphate aldolase homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000

A:Accession: A64477

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,

Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kane, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64477
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-181 <BUL>
A:Cross-references: GB:U67582; GB:I77117; NID:gl592064; PIDN:AAB99428.1; PID:gl592067; T
C:Genetics:
A:Map position: FOR1381152-1381697
C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match 11.5%; Score 152; DB 1; Length 181;
Best Local Similarity 24.1%; Pred. No. 2.8e-06;
Matches 42; Conservative 38; Mismatches 66; Indels 28; Gaps 5;

QY 29 ELCKQFVHLGWVTGGGSLKHGNEIYAPSGVQKRIQPEDMFVCDINEQDISGPPAS 88
DB 8 KICKRLYDRKVVGGGNSVKGDKIYLTPTGILGFLKEDDAEMDLGDNVIGKPTS 67
QY 89 KKLKKSQCTPLFMNAYTWRG-AGAVIHTHSCAAVWATLLFPQGFKITHQEMIKGIRKCT 147
DB 68 EK-----NLHLMYRKNDINAIHTHSLISTFLSTI--NKBIELTTPGKIFLKKI- 117
QY 148 SGGYRYDDMLVPIENTPEKDKERMAHMEYPPSCAVLVRHGVVYVGE 201
DB 118 --GYVDYEEAGSLAEETAKR-----DEDVILLKNGHGVCLGK 154

RESULT 9
A:72396
L-fuculose-phosphate aldolase homolog - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: A72396
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: A72396
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <ARN>
A:Cross-references: GB:AE001710; GB:AE000512; NID:94980775; PIDN:AAD35371.1; PID:9498078
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TMO283
C:Superfamily: L-ribulose-phosphate 4-epimerase

Query Match 11.3%; Score 148.5; DB 1; Length 254;
Best Local Similarity 26.2%; Pred. No. 8.6e-06;
Matches 64; Conservative 44; Mismatches 95; Indels 41; Gaps 13;

QY 4 COAQGCCSRPCGAQ---DKHPRFLIPELCKQFHLGWVTGGGSLKHGNEIYAP 60
DB 27 CVDNFDCFTTHGKGSMEYKERKELYNALLLEKY--GLVAYTSGNVSVRIGDHVLIKPS 84
QY 61 GVOKERIQPEDMFVCDINEQDISGPPASKKLSQCTPLFMNAYTWRGAGAVIHTHSCAA 120
DB 85 GVPYTELKPEDFVVDLEGNVIEG---EKKPSVDYATHLYLKH-LDWAKSVIHTHTFA 140
QY 121 VMATLLFPQGFKITHQEMIKGIRKCTSGYRYDDML--VVPITENTPEEKD-LKRNA 177
DB 141 MWAIL-----EKSIPVL--CTA-----HADVFEEIPLTEYAPVGSEAGKAVV 183
QY 178 HAMNEYPPSCAVLVRHGVVWGETWEKA--KTMCEYDYLFDIAVS--MKWGLDPTOL 233
DB 184 KVIGK---SGAVLLRKHGMVIGVTSVDDAVKAI-----FLEEVAKAAYFATLAKPTPL 235

QY 234 PVGE 237
DB 236 PPDE 239

RESULT 10
A:3533
L-fuculose phosphate aldolase (EC 4.1.2.17) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AC3533
R:DeIvecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.
; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis.
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3533
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53430.1; PID:gl7984327; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10189
A:Map position: II
C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 10.6%; Score 140; DB 2; Length 244;
Best Local Similarity 23.0%; Pred. No. 4.8e-05;
Matches 52; Conservative 33; Mismatches 75; Indels 66; Gaps 10;

QY 4 COAQGCCSRPCGAQDKHPRFLIPELCKQFHLGWVTGGGSLKHGNEIYAPSGVQ 63
DB 24 CKQMNCDSALLARQS-----IVDAMRSPKGFHNGSSGNISVREGGHIWVTPGA- 75
QY 64 KERIQPEDMFVCDINEQDISGPPASKKLSQCTPLFMNAYTWRG---AGAVIHTHSCAA 120
DB 76 TSTMDPQDMSLVSEGLHAGKIPSEWR-----IHTIMEAHPEAGAVVISHADAC 127
QY 121 VMATLL-----FPQGFKITHQEMIKGIRKCTSGGYRY--DDMLVVPPIEN 165
DB 128 VALSCLRKPLPPFHYMTASFGSEV-----PCAS---YRVFGSDALAYEVV-- 170
QY 166 TPPEKDKERMAHMEYPPSCAVLVRHGVVWGETWEKAKTWCE 211
DB 171 -----RAMGHR---ACIMASHGMVWVGRDLAHLALAE 201

RESULT 11
A:7804
L-ribulose-phosphate 4-epimerase homolog (EC 5.1.3.-) yias - Escherichia coli (strain K-1)
N:Alternate names: hypothetical protein o231
C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 01-Mar-2002
C:Accession: S47804; A65158
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47804
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <PLU>
A:Cross-references: EMBL:U00039; NID:946582; PIDN:AAB18560.1; PID:9466721
A:Experimental source: strain K-12, substrain MG1655
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A65158
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-231 <BLAT>

RESULT 13

AI0977
probable sugar isomerase (EC 5.1.3.4) [imported] - Salmonella enterica subsp. enterica serovar typhi
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0977
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, R.; S.; Moulle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar AB0502; MUID:21534947; PMID:11677608
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0977
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07949.1; PID:gi6504938; GSPDB:GN00176
C:Genetics:
C:Superfamily: L-ribulose-phosphate 4-epimerase
C:Keywords: isomerase

Query Match 10.0%; Score 132; DB 2; Length 231;
Best Local Similarity 28.5%; Pred. No. 0.00023;
Matches 61; Conservative 23; Mismatches 94; Indels 36; Gaps 11;

QY 38 GWVTGTGGGSLKHEGNE--IVIAPSGVOKERIQPEDMFVCDINEQDISGPPASKLKKSQ 95
DB 20 GLVFTTGWNSAVDETRKLMWIKPSGVEYVWTDMMVVEI-----ASGVVEGNKKPSS 75
QY 96 CTPLFMNAY-TMRGAGAVIHTSKAAVN---ATLFPGQGEFKITHOEMIKIRKCTS--- 148
DB 76 DTATHALYRRYPQIGGVVTHSRHATWSQAGLDLPA-WGTTHADYFVGAIPTRLMT 133
QY 149 ----GGYRYDDMLVPIENTPEKDLKERMAHANEYPDSCAVLRHGVVVG---- 200
DB 134 VEEINGEYEQ---TGEVVIITFEERGLDPA-----QIP---AVLVHSHGPPFANGKNA 181
QY 201 ETWEKAKTMCEC-YDYLFDIAVSMKMGWLDPTQL 233
DB 182 DAVNNVVEECATWGLFSRQLAPQIPDMQPELL 215

RESULT 14

S73671
L-ribulose-phosphate 4-epimerase (EC 5.1.3.4) arab - Mycoplasma pneumoniae (strain ATCC 29216)
N:Alternate names: hypothetical protein P02_orf242
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29242
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73671
R:Hammelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73671
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-242 <HIN>
A:Cross-references: ENBL:AE000032; GB:U00089; NID:gi674011; PIDN:AAB95992.1; PID:gi674022
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: arab
A:Genetic code: SGC3
C:Superfamily: L-ribulose-phosphate 4-epimerase
C:Keywords: isomerase; zinc
F:80,99,101,175/Binding site: zinc (Asp, His, His, His) #status predicted

Query Match 10.0%; Score 132; DB 2; Length 242;
Best Local Similarity 27.9%; Pred. No. 0.00025;
Matches 55; Conservative 25; Mismatches 73; Indels 44; Gaps 10;

Search completed: June 1, 2004, 13:35:23
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2004, 13:29:16 ; Search time 46 Seconds
(without alignments)
1653.041 Million cell updates/sec

Title: US-09-937-905-2

Perfect score: 1317

Sequence: 1 MSGCQAQDCCSRFCGAQDK.....SMKXMGIDPTQLPVGENGIV 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1317	100.0	241	11 Q9WVQ5	Q9WVQ5 mus musculus
2	1307	99.2	241	11 Q8BP46	Q8BP46 mus musculus
3	1246.5	94.6	242	4 Q96GX9	Q96GX9 homo sapien
4	1239.5	94.1	242	4 Q9V318	Q9V318 homo sapien
5	1235.5	93.8	242	4 Q8WVU2	Q8WVU2 homo sapien
6	1234.5	93.7	242	4 Q96HK2	Q96HK2 homo sapien
7	786	59.7	227	5 Q9VTF93	Q9VTF93 drosophila
8	772.5	58.7	153	4 Q9H528	Q9H528 homo sapien
9	540	41.0	507	10 Q9FN41	Q9FN41 arabidopsis
10	433.5	32.9	284	5 Q23261	Q23261 caenorhabdi
11	417	31.7	221	3 Q9HE08	Q9HE08 schizosacch
12	289	21.9	240	5 Q8TR31	Q8TR31 heterodera
13	274.5	20.8	192	3 Q9UT22	Q9UT22 schizosacch
14	219.5	16.7	209	16 Q31668	Q31668 bacillus su
15	211	16.0	207	16 Q7V8Y6	Q7V8Y6 prochloroco
16	203	15.4	202	16 Q828L0	Q828L0 streptomyce

17	202	15.3	212	16	Q81956	Q81956 bacillus ce
18	201.5	15.3	211	16	Q7U4V0	Q7U4V0 synechococc
19	201	15.3	212	16	Q81MJ0	Q81MJ0 bacillus an
20	195	14.8	204	16	Q884P3	Q884P3 pseudomonas
21	179.5	13.6	249	16	Q8EXC1	Q8EXC1 leptospira
22	166	12.6	217	16	Q8P9N3	Q8P9N3 xanthomonas
23	165	12.5	217	16	Q8PLG0	Q8PLG0 xanthomonas
24	164	12.5	205	16	Q91342	Q91342 pseudomonas
25	161.5	12.3	241	16	Q829J0	Q829J0 streptomyce
26	155	11.8	218	16	Q87C38	Q87C38 xylella fas
27	153	11.6	218	16	Q9PBD5	Q9PBD5 xylella fas
28	148.5	11.3	254	16	Q9WYB9	Q9WYB9 thermotoga
29	146	11.1	231	16	Q8FC7	Q8FC7 escherichia
30	145	11.0	430	16	Q7UYE2	Q7UYE2 rhodospirillum
31	144	10.9	190	17	Q8TV16	Q8TV16 methanopyru
32	140	10.6	219	16	Q83HE2	Q83HE2 tropheryma
33	140	10.6	244	16	Q8YDI7	Q8YDI7 brucella me
34	139.5	10.6	240	16	Q9RJ24	Q9RJ24 streptomyce
35	138.5	10.5	191	17	Q27457	Q27457 methanobact
36	136	10.3	226	16	Q8PEU7	Q8PEU7 xanthomonas
37	134.5	10.2	219	2	Q8VQM1	Q8VQM1 brucella ab
38	132	10.0	231	16	Q8Z2D2	Q8Z2D2 salmonella
39	132	10.0	243	16	Q8EPX0	Q8EPX0 mycoplasma
40	131.5	10.0	219	16	Q8FVU6	Q8FVU6 brucella su
41	131	9.9	231	16	Q8ZL75	Q8ZL75 salmonella
42	131	9.9	238	2	Q93Q57	Q93Q57 klebsiella
43	130.5	9.9	236	16	Q93DA6	Q93DA6 streptococc
44	130	9.9	227	16	Q97NJ4	Q97NJ4 streptococc
45	128	9.7	223	16	Q83FR4	Q83FR4 tropheryma

ALIGNMENTS

RESULT 1

Q9WVQ5 PRELIMINARY; PRT; 241 AA.
AC Q9WVQ5; DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE MMRP19 (Monocyte macrophage 19).
GN MMRP19 OR MMR19.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Sha S., Aoki Y., Nishi Y.;
RT "A cDNA sequence from murine monocyte-macrophage."
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary Gland;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB028863; BAA78906.1; -
DR EMBL; BC028434; AAH28434.1; -
DR MGD; MGI:1926788; Mmrp19.
DR InterPro; IPR001303; Aldolase_II_N.
DR Pfam; PF00596; Aldolase II; 1.
SQ SEQUENCE 241 AA; 4366CF4AD2239DB8 CRC64;

Query Match 100.0%; Score 1317; DB 11; Length 241;
Best Local Similarity 100.0%; Pred. No. 2.3e-118;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSGCQAQDCCSRFCGAQDKHEPRFLIPELCKQFVHLGWVTCTGGGSLKHGNEIYIAPS 60
DB 1 MSGCQAQDCCSRFCGAQDKHEPRFLIPELCKQFVHLGWVTCTGGGSLKHGNEIYIAPS 60
QY 61 GYQKERIQPEDMFVCDINEQDISGPPASKKLKKSQCTPLFMNAYTWRGAGAVIHTHSKAA 120

Db 1 MSGCDAGDCCRRGCAQDKBHPRLIPELCKQFYHLGVTGTGGISLKHGDEIYIAP 60
 QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKA 119
 Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKA 120
 QY 120 AVMATLLFPQGFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
 Db 121 AVMATLLFPQGFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180
 QY 180 MNEYPDSCAVLVRHGVYVGTWEKAKTMCCEYDYLFDIAVSMKKMGDPTQLPVGNG 239
 Db 181 MNEYPDSCAVLVRHGVYVGTWEKAKTMCCEYDYLFDIAVSMKKMGDPTQLPVGNG 240
 QY 240 IV 241
 Db 241 IV 242

RESULT 5

Q8WU2 PRELIMINARY; PRT; 242 AA.
 AC Q8WU2;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC017594; AAH17594.1; -.
 DR InterPro; IPR001303; Aldolase_II N.
 DR Pfam; PF00596; Aldolase_II; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 242 AA; 27215 MW; F8C4A371E3C63FC2 CRC64;

Query Match 93.8%; Score 1235.5; DB 4; Length 242;
 Best Local Similarity 93.8%; Pred. No. 1.6e-110;
 Matches 227; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 1 MSGCQA-QGDCSRPCGAQDKBHPRLIPELCKQFYHLGVTGTGGISLKHGDEIYIAP 59
 Db 1 MSGCDAGDCCRRGCAQDKBHPRLIPELCKQFYHLGVTGTGGISLKHGDEIYIAP 60
 QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKA 119
 Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKA 120
 QY 120 AVMATLLFPQGFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
 Db 121 AVMATLLFPQGFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180
 QY 180 MNEYPDSCAVLVRHGVYVGTWEKAKTMCCEYDYLFDIAVSMKKMGDPTQLPVGNG 239
 Db 181 MNEYPDSCAVLVRHGVYVGTWEKAKTMCCEYDYLFDIAVSMKKMGDPTQLPVGNG 240
 QY 240 IV 241
 Db 241 IV 242

RESULT 6

Q96HK2 PRELIMINARY; PRT; 242 AA.
 ID Q96HK2
 AC Q96HK2
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Similar to CGI-29 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC008440; AAH08440.1; -.
 DR InterPro; IPR001303; Aldolase_II N.
 DR Pfam; PF00596; Aldolase_II; 1.
 SQ SEQUENCE 242 AA; 27112 MW; 691C60447B723C1B CRC64;
 Query Match 93.7%; Score 1234.5; DB 4; Length 242;
 Best Local Similarity 93.4%; Pred. No. 2e-110; 5; Indels 1; Gaps 1;
 Matches 226; Conservative 10; Mismatches 5;
 QY 1 MSGCQA-QGDCSRPCGAQDKBHPRLIPELCKQFYHLGVTGTGGISLKHGDEIYIAP 59
 Db 1 MSGCDAGDCCRRGCAQDKBHPRLIPELCKQFYHLGVTGTGGISLKHGDEIYIAP 60
 QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKA 119
 Db 61 SGVQKERIQPEDMFVCDINEQDISGPPASKLKKSOCTPLFNAYTMRGAGAVIHTSKA 120
 QY 120 AVMATLLFPQGFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 179
 Db 121 AVMATLLFPQGFKITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPEEKDKERMAHA 180
 QY 180 MNEYPDSCAVLVRHGVYVGTWEKAKTMCCEYDYLFDIAVSMKKMGDPTQLPVGNG 239
 Db 181 VNEYPDSCAVLVRHGVYVGTWEKAKTMCCEYDYLFDIAVSMKKMGDPTQLPVGNG 240
 QY 240 IV 241
 Db 241 IV 242

RESULT 7

Q9VY93 PRELIMINARY; PRT; 227 AA.
 ID Q9VY93;
 AC Q9VY93;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE CG11134 protein (RE61999p).
 GN CG11134.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Rhue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Berkeley.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Man K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB003493; AAF48310.1; -
 DR EMBL; AY071553; AAL49175.1; -
 DR FlyBase; FBgn0030518; CG11134.
 DR InterPro; IPR001303; Aldolase II_N.
 DR Pfam; PF00596; Aldolase II; 1.
 SQ SEQUENCE 227 AA; 26011 MW; 7F2E505906CE155D CRC64;

Query Match 59.7%; Score 786; DB 5; Length 227;
 Best Local Similarity 70.5%; Pred. No. 2.4e-67;
 Matches 148; Conservative 20; Mismatches 42; Indels 0; Gaps 0;

QY 21 EHPRLPELCKQFVHLGWVTGTGGISLKHGNEIYAPSGVQKRIQPEDMFVCDINEQ 80
 Db 12 EHPRLPELCKQFVHLGWVTGTGGISLKHGNEIYAPSGVQKRIQPEDMFVCDITGK 71

QY 81 DISGPPASKLKSQCTPLFMNAYTMRGAGAVIHTHSAKAVMATLLPFGQBFKITHQEMI 140
 Db 72 DLQLPPEIKGLKKSQCTPLFMAYQHRQAGAVIHTHSAKAVMATLLPFGQBFKITHQEMI 131

QY 141 KGIRKCTGGYRYDMLVPIENTPEKDKERMAHMEYPPSCAVLVRHGVYWG 200
 Db 132 KGVIDEADKRLRYDEELVPIENTPTPEKDLADSYAAMMEYPPGCSAILVRHGVYWG 191

QY 201 ETWEKAKTMCYDYLFDAVSMKMGDLDP 230
 Db 192 QNWEKAKTMCYDYLFDAVSMKMGDLDP 221

RESULT 8
 Q9H528 PRELIMINARY; PRT; 153 AA.

AC Q9H528;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE DJ179L10.2 (Similar to CGI-29 protein) (Fragment).
 GN DJ179L10.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Moore M.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL138810; CAC12642.1; -
 DR InterPro; IPR001303; Aldolase II_N.
 DR Pfam; PF00596; Aldolase II; 1.
 FT NON TER 153
 SQ SEQUENCE 153 AA; 16918 MW; D88A0ADE331E12EA CRC64;

Query Match 58.7%; Score 772.5; DB 4; Length 153;
 Best Local Similarity 92.8%; Pred. No. 2.9e-66;
 Matches 142; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 MSGCOA-QGCCSRPCGAQKHEPRFLIPELCKQFVHLGWVTGTGGISLKHGNEIYAP 59
 Db 1 MSGCDAREGDCSRRCGAQKHEPRYLIPELCKQFVHLGWVTGTGGISLKHGDEIYAP 60

QY 60 SGVQKERIQPEDMFVCDINEQDISGPPASKLKSQCTPLFMNAYTMRGAGAVIHTHSA 119
 Db 61 SGVQKERIQPEDMFVCDINEKDISGPPSKLKSQCTPLFMNAYTMRGAGAVIHTHSA 120

QY 120 AVMATLLPFGQBFKITHQEMIKGIRKCTSG3YY 152
 Db 121 AVMATLLPFGQBFKITHQEMIKGIRKCTSG3YY 153

RESULT 9
 Q9FN41 PRELIMINARY; PRT; 507 AA.

ID Q9FN41;
 AC Q9FN41;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Similarity to enolase-phosphatase.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98162728; PubMed=9501997;
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
 RT Sequence features of the regions of 1,191,918 bp covered by seventeen
 RT physically assigned P1 clones.";
 RL DNA Res. 4:401-414(1997).
 DR EMBL; AB007644; BAB10715.1; -
 DR GO; GO:0016787; F-hydrolase activity; IEA.
 DR GO; GO:0008152; P-metabolism; IEA.
 DR InterPro; IPR001303; Aldolase II_N.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00596; Aldolase II; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 SQ SEQUENCE 507 AA; 56520 MW; 2F99D9F52FA1D5DC CRC64;

Query Match 41.0%; Score 540; DB 10; Length 507;
 Best Local Similarity 49.6%; Pred. No. 2.8e-43;
 Matches 112; Conservative 32; Mismatches 60; Indels 22; Gaps 7;

QY 26 LIPELCKQFVHLGWVTGTGGISLKHGNEIYAPSGVQKRIQPEDMFVCDI 77
 Db 27 LVTELCHRYTQGVWVGTTGGISITMKVHDASIPKEQLIVMSFGVQKERMQEDMYILSA 86

QY 78 NEQDISGPP-----ASKKLKSQCTPLFMNAYTMRGAGAVIHTHSAKAVMATLLP-GEFF 132
 Db 87 NGSIISTSPKPYNPKPKCTDCAPLFMKAYEMRNAGAVIHTHSGMESCLVTMLNPQAEF 146

```

OC 133 KITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 192
OX 147 RITHMEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 200
RN 147 RITHMEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 200
RP 147 RITHMEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 200
RC STRAIN=972h-;
RA Seger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL512487; CAC19735.1; -
DR GeneDB SPombe; SPAC20H4.05C; -
DR InterPro; IPR001303; Aldolase II_N.
DR Pfam; PF00596; Aldolase II; 1.
DR SQU SEQUENCE 221 AA; 25195 MW; B959011B581EAB74 CRC64;

Query Match 31.7%; Score 417; DB 3; Length 221;
Best Local Similarity 39.6%; Pred. No. 6.8e-32;
Matches 90; Conservative 40; Mismatches 75; Indels 22; Gaps 7;

QY 3 GCOAGGCCSRPCGAQDKHEPRFLPELCKQFYHLGWTGTGGGSLKHGNEIYIAPSGV 62
DB 10 GCLSGDL--KKCGE-----LICEICRDLYTSGWVTGT-----GDAIVIAPSGV 51
QY 63 QKRIQPEDMFVCDINEQDISGPPASKLKKSCQCTPLFMNAYT-MRGAGAVIHTSKAAV 121
DB 52 QKERMELHFLFVMSLITREYMPA-LRLKPSQCTPLFLAVVTSLRDAYACIHTSQEAI 110
QY 122 MATLFPQOE-FKITHQEMIKGIRK-CTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHA 179
DB 111 LLSTLTFADSDHFSATGFEVLSPKSGKNGPHKPTDKIKIPFINNTAHESDLHDSLOEA 170
QY 183 MNEYPDSCAVLVRHGVVWGETWEKATCMCEYDYLFDIAVSMKKM 226
DB 171 INLYPDTCAVIRDHGYCWGTQDWTQNTKMTAEVFLFOAYLRRRL 217

RESULT 12
QTA31 PRELIMINARY; PRT; 240 AA.
AC Q8TA31;
DC 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Heterodera glycines (soybean cyst nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
OC NCBI_TaxID=51029;
RN [1]_
RP SEQUENCE FROM N.A.
RA Gao B., Allen R., Baum T.J., Davis E.L., Hussey R.S.;
RT "Cloning putative parasitism genes from Heterodera glycines."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF345795; AAL78223.1; -
DR GO; GO:0005874; C:microtubule; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007018; P:microtubule-based movement; IEA.
DR InterPro; IPR001303; Aldolase II_N.
DR InterPro; IPR002453; Beta tubulin.
DR Pfam; PF00596; Aldolase II; 1.
DR PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 240
SQ SEQUENCE 240 AA; 26062 MW; 77862AEE08142DBA CRC64;

Query Match 21.9%; Score 289; DB 5; Length 240;
Best Local Similarity 35.1%; Pred. No. 1.5e-19;
Matches 67; Conservative 37; Mismatches 77; Indels 10; Gaps 4;

QY 27 IPELCKQFYHLGWTGTGGGSLKHGNEIYIAPSGYKQRIQPEDMFVCDINEQDISGPP 86
DB 17 LAELIHFALGWRDNGGMAVLNGAVFGSPTSQKEKVPENDLFDVDTATGTVLKRP 76
QY 87 ASKLLKKSQCTPLFMNAYTMRGAGAVIHTSKAAVMAVATLFPQGEFKITHQEMIKGIRK 146
```

```

QY 133 KITHQEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 192
DB 147 RITHMEMIKGIRKCTSGGYRYDDMLVPIIENTPPEEKDKLKERMAHAMNEYDPDSCAVLVR 200
QY 193 RHGVVWGETWEKATCMCEYDYLFDIAVSMKKGLD---FTQLPV 235
DB 201 NHGVYWGDSWTHAKTQAEVCYHYLFDAAIKHQLGLDAATPDHGPI 246

RESULT 10
Q23261 PRELIMINARY; PRT; 284 AA.
AC Q23261;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ZC373.5 protein.
GN ZC373.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]_
RP SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z49131; CAA8977.1; -
DR FIR; T27523; T27523.
DR WormPep; ZC373.5; CE02378.
DR InterPro; IPR001303; Aldolase II_N.
DR Pfam; PF00596; Aldolase II; 1.
DR SQU SEQUENCE 284 AA; 32367 MW; 4DC5B94439F2E529 CRC64;

Query Match 32.9%; Score 433.5; DB 5; Length 284;
Best Local Similarity 39.9%; Pred. No. 2.4e-33;
Matches 87; Conservative 35; Mismatches 75; Indels 21; Gaps 1;

QY 29 ELCKQFYHLGWTGTGGGSLKHGNEIYIAPSGYKQRIQPEDMFVCDINEQDISGPPAS 88
DB 35 ELMIQYKLGWRGSGGAGCTSGSLMSPALQKRIREQDVFYNNKDKTEVQRPN 94
QY 89 KKLKKSQCTPLFMNAYTMRGAGAVIHTSKAAVMAVATLFPQGEFKITHQEMIKGIRKCTS 148
DB 95 KRITVSSCSVLFLIMKETGSECVIHTSKCANLITOLIKSNVFEISHQEIYKGIYDPFS 154
QY 149 GGYRYDDMLVPIIENTPPEEKDL-----KERMAHAMNEYDPDSC 187
DB 155 GKALKYSDTLTPIIDNMPSESQLLVCAINRSPMQDPCFNIALFOEIRGVLENTPOAI 214
QY 188 AVLVRHGVVWGETWEKATCMCEYDYLFDIAVSMKK 225
DB 215 AVLVRHGVVWGETWEKATCMCEYDYLFDIAVSMKK 252

RESULT 11
Q9HE08 PRELIMINARY; PRT; 221 AA.
AC Q9HE08;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative phosphate epimerase.
GN SPAC20H4.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
```

Db 77 QNAAVSPTSATCGLLMTNT-GLNCVITHSKYANLVSQLVITGNEFAIQNQMIGQYENR 132
Qy 147 TSGGYRYD-DMLAVPIIENTPKDKERMAHMEYPDSCAVLVRHGVYVWGGETWE 204
Db 133 SSG--LRDNDVRUVFIVDSNELNEQMSPLVRLTLDKYTEASALLVRGFFVFGSSLA 190
Qy 205 KA--KTMCECY 213
Db 191 GSLPKWMLF 201
RESULT 13
Q9UT22 PRELIMINARY; PRT; 192 AA.
AC Q9UT22;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN SPAC9.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h;
RA Wedler H., Dusterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121764; CAB57424.1; -;
DR PIR; T39191; T39191.
DR GeneDB SPombe; SPAC9.06C; -;
DR InterPro; IPR001303; Aldolase_II_N.
DR Pfam; PF00596; Aldolase_II; 1.
KW Hypothetical protein.
SQ SEQUENCE 192 AA; 21797 MW; F988CD3983007191 CRC64;
Query Match 20.8%; Score 274.5; DB 3; Length 192;
Best Local Similarity 34.6%; Pred. No. 2.8e-18;
Matches 71; Conservative 36; Mismatches 73; Indels 25; Gaps 7;
Qy 26 LIPELCKQFYHLGWVT-GTGGISLKHGNEIYAPSGVQKERIQPDMFVCDINEQDISG 84
Db 9 LIELIPIHYSLGWKKFGSGYALCVK-----DRVQRDFITENDIVTFLNSQSVT- 58
Qy 85 PPASKKLKKSQCTPLFMNAYTMRGAGAV--IHTHKAAMVATLLFPQGEFKITHQEMING 142
Db 59 -----KDLVNWAYIFSWLSNMDAVACIYSTSVAAVGASMY--NEKFTTQSKEMING 108
Qy 143 IRKCT-SGGYRYDDMLVPIIENTPKDKERMAHMEYPDSCAVLVRHGVYVWGE 201
Db 109 IPKGNPSAGYLCFFDTLEVPILHN-GDSKTIIDELKKVIELPQICAVLIRHGVIWGA 167
Qy 202 TWEKAKTMCECYDYLFDIAVSMKKM 226
Db 168 TWEKAKTMCECYEYLFELDYKLKTL 192
RESULT 14
O31668 PRELIMINARY; PRT; 209 AA.
AC O31668;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE YKRY protein.
GN YKRY.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Avevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoeft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Rapoport G., Rey M., Reynolds S.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rose M., Sadale Y.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "the complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99111; CAB13234.1; -;
DR PIR; A69864; A69864.
DR InterPro; IPR001303; Aldolase_II_N.
DR Pfam; PF00596; Aldolase_II; 1.
KW Complete proteome.
SQ SEQUENCE 209 AA; 23489 MW; D4E8BE3544026FAB CRC64;
Query Match 16.7%; Score 219.5; DB 16; Length 209;
Best Local Similarity 30.9%; Pred. No. 6.1e-13;
Matches 64; Conservative 34; Mismatches 88; Indels 21; Gaps 7;
Qy 27 IPLECKQFYHLGWVTGTGGISLKHGNE--IYIAPSGVQKERIQPDMFVCDINEQDIS 83
Db 11 LAEVRKELAEKRDWFPATSGNISIKVTDEPLFLVTASGDKRKETVEDFLVDQN----- 65
Qy 84 PPASK--KLKKSQCTPLFMNAYTMRGAGAVIHTHKAAMVATLLFPQGEFKITH--QEM 139
Db 66 GEPAESGHSKPSAETLLHLYNKTNAGCCLHVTNNVISELY-GDQKIKTPGQEI 124
Qy 140 IKGIRKCTSGGYRYDDMLVPIIENTPKDKERMAHMEYPDSCAVLVRHGVYVW 199
Db 125 IKAL-----GLWEENAEVTVPIIENPAHIPTLAALFAEEISE--DSGAVLIRNHGITAW 176
Qy 200 GETWEKAKTMCECYDYLFDIAVSMKKM 226
Db 177 GKTAFAEKRVLEAYEFLFSYHLKLT 203
RESULT 15
Q7V8Y6 PRELIMINARY; PRT; 207 AA.
ID Q7V8Y6;
AC Q7V8Y6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

